



Πανεπιστήμιο Δυτικής Αττικής

Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

Πρόγραμμα Μεταπτυχιακών Σπουδών: ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ

**Αλγοριθμική λήψη αποφάσεων, Διακρίσεις και
Ανθρώπινα Δικαιώματα**

Γεώργιος Παπακωνσταντίνου

A.M: CSCYB2021

Διπλωματική Εργασία

Επιβλέπων: Δρ. Στέφανος Γκρίτζαλης

Αιγάλεω, Φεβρουάριος 2023

Copyright© Γεώργιος Παπακωνσταντίνου, 2023

All rights reserved. Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας διπλωματικής εργασίας εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν την χρήση της διπλωματικής εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Η έγκριση της διπλωματικής εργασίας από το Πανεπιστήμιο Δυτικής Αττικής δεν δηλώνει αποδοχή των γνώμων του συγγραφέα.



Πανεπιστήμιο Δυτικής Αττικής

Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

Πρόγραμμα Μεταπτυχιακών Σπουδών: ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ

Αλγοριθμική λήψη αποφάσεων, Διακρίσεις και Ανθρώπινα Δικαιώματα

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η μεταπτυχιακή διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι
Εξεταστική Επιτροπή:

Α/Α	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Στέφανος Γκρίτζαλης	Καθηγητής Τμήμα Ψηφιακών Συστημάτων, Πανεπιστήμιο Πειραιώς / Εισηγητής - Επιβλέπων	
2	Παναγιώτης Γιαννακόπουλος	Καθηγητής Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών, Πανεπιστήμιο Δυτικής Αττικής / Μέλος Εξεταστικής Επιτροπής	
3	Κωνσταντίνος Μαυρομάτης	Λέκτορας Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών, Πανεπιστήμιο Δυτικής Αττικής / Μέλος Εξεταστικής Επιτροπής	

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος **Γεώργιος Παπακωνσταντίνου** του **Φραγκίσκου**, με αριθμό μητρώου **escyb2021** φοιτητής του Προγράμματος Μεταπτυχιακών Σπουδών «**ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ**» του Τμήματος **Μηχανικών Πληροφορικής και Υπολογιστών** της Σχολής **Μηχανικών** του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



Περίληψη

Η διείσδυση της αλγοριθμικής λήψης αποφάσεων σε πολλούς τομείς της σύγχρονης ζωής, όπως η ιατρική, η δικαιοσύνη, η ασφάλεια, η εκπαίδευση κ.α. κομίζει πολλαπλά οφέλη, μεταξύ των οποίων ξεχωρίζει η μείωση των ανθρώπινων λαθών κατά τη λήψη αποφάσεων, η καταπολέμηση της μεροληψίας κατά τη λήψη αποφάσεων από ανθρώπους, η βελτίωση της ακρίβειας των λαμβανομένων αποφάσεων, η επιτάχυνση λήψης απόφασης κ.α.. Ωστόσο, λαμβάνοντας υπόψη το γεγονός ότι όταν οι αποφάσεις λαμβάνονται από ανθρώπους δεν επενεργεί μόνον η πιθανή μεροληψία των ίδιων των ανθρώπων αλλά και ένα σύνολο νόμων, κανονισμών και ηθικών περιορισμών ή αρχών, εγείρονται σημαντικά ζητήματα και ανησυχία καθώς τα υπολογιστικά συστήματα είναι δύσκολο να λάβουν μια απόφαση εφαρμόζοντας το ίδιο σύνολο νόμων κανονισμών και ηθικών περιορισμών ή αρχών. Εντούτοις, μολονότι μεγάλη μερίδα των ακαδημαϊκών θεωρεί πως η αλγόριθμοι και λήψη αποφάσεων μπορεί να οδηγήσει σε μείωση της μεροληψίας σε πολλές εκφάνσεις της σύγχρονης ζωής, υπάρχει σημαντικός αντίλογος και ανησυχία αναφορικά με την πιθανή παραβίαση των ανθρωπίνων δικαιωμάτων και την πιθανή δημιουργία καλλιέργεια ή διεύρυνση διακρίσεων. Στα πλαίσια της παρούσης πτυχιακής εργασίας εξετάστηκε η αλγοριθμική λήψη αποφάσεων τόσο υπό το πρίσμα των πλεονεκτημάτων και μειονεκτημάτων εφαρμογής της όσο και υπό το πρίσμα του πεδίου εφαρμογής της, με έμφαση στο σεβασμό των ανθρωπίνων δικαιωμάτων και στην αποφυγή δημιουργίας διακρίσεων. διαφαίνεται ότι η αυστηρή μεταφορά του νομοθετικού πλαισίου στους ίδιους τους αλγορίθμους μέσω των οποίων λαμβάνονται οι αποφάσεις μπορεί να οδηγήσει σε σχεδόν εξάλειψη της πιθανότητας δημιουργίας διακρίσεων και παραβίασης των ανθρωπίνων δικαιωμάτων. ωστόσο εγείρονται ανησυχίες όσον αφορά στην επίδραση που μπορεί να έχει η αλγοριθμική λήψη αποφάσεων με επίγνωση των διακρίσεων στο κόστος της διαδικασίας αλγοριθμικής λήψης αποφάσεων και εν γένει στην ίδια την εφικτότητα και βιωσιμότητα εφαρμογής της. Μέσω της παρούσης πτυχιακής εργασίας προτείνεται ότι τα οφέλη από την αλγοριθμική λήψη αποφάσεων υπερκερνούν τα μειονεκτήματα και τους φόβους περί δημιουργίας διακρίσεων και προτείνεται η συνέχιση και εντατικοποίηση της έρευνας προκειμένου να δημιουργηθεί ένα σύνολο αρχών επί των οποίων θα σχεδιάζονται οι αλγόριθμοι προκειμένου να παραμένουν βιώσιμοι αλλά και να αποκλείεται σε μεγάλο ποσοστό η δημιουργία διακρίσεων.

Abstract

The penetration of algorithmic decision-making in many areas of modern life, such as medicine, justice, security, education, etc. brings multiple benefits, among which stands out the reduction of human errors in decision-making, the fight against bias in decision-making by humans, the improvement of the accuracy of decisions made, the acceleration of decision making, etc. However, taking into account the fact that when decisions are made by humans it is not only the possible bias of the people themselves that acts, but also a set of laws, regulations and ethical restrictions or principles, important issues and concerns arise as computer systems are difficult to make a decision applying the same set of laws, regulations and ethical restrictions or principles. However, although a large proportion of academics believe that algorithms and decision-making can lead to a reduction in bias in many aspects of modern life, there is considerable controversy and concern regarding the possible violation of human rights and the possible cultivation or widening of discrimination. In the context of this thesis, algorithmic decision-making was examined both in the light of the advantages and disadvantages of its implementation and in the light of its scope, with emphasis on respect for human rights and non-discrimination. It appears that the strict transposition of the legislative framework into the very algorithms through which decisions are taken may lead to the almost elimination of the possibility of discrimination and violation of human rights. However, there are concerns about the impact that discrimination-aware algorithmic decision-making can have on the cost of algorithmic decision-making and in general on the feasibility and sustainability of its implementation. Through this thesis, it is proposed that the benefits of algorithmic decision-making outweigh the disadvantages and fears of discrimination and it is proposed that the research should continue and be intensified in order to create a set of principles on which algorithms will be designed in order to remain viable but also to exclude to a large extent the creation of discrimination.

Λίστα εικόνων

Εικόνα 1: Αλγοριθμικό σύστημα το οποίο χρησιμοποιεί τεχνικές μάθησης μηχανών	21
Εικόνα 2: Τυπικό δέντρο αποφάσεων για την ταξινόμηση δεδομένων (Dobra, 2009)	34
Εικόνα 3: Πληροφοριακά συστήματα – μοντέλο πυραμίδας τριών επιπέδων (Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017)	37
Εικόνα 4: Πληροφοριακά συστήματα – μοντέλο πυραμίδας πέντε επιπέδων (Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017)	38
Εικόνα 5: Πληροφοριακά συστήματα – μοντέλο πυραμίδας πέντε επιπέδων (Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017)	39
Εικόνα 6: Τυπική δομή αλγοριθμικής λήψης αποφάσεων με γνωστά τα δημογραφικά και άλλα προστατευόμενα χαρακτηριστικά (Žliobaitė, 2017a)	45
Εικόνα 7: Αλληλεπίδραση γνώσεων, τεχνικών και εργαλείων για τη δίκαια αλγοριθμική λήψη αποφάσεων (Žliobaitė, 2017b)	47

Περιεχόμενα

Περίληψη	5
Abstract.....	6
Λίστα εικόνων.....	7
Περιεχόμενα.....	8
Κεφάλαιο 1 ^ο : Εισαγωγή	10
1.1 Αλγοριθμική λήψη αποφάσεων στη σύγχρονη κοινωνία.....	10
1.2 Διακρίσεις στην αλγοριθμική λήψη αποφάσεων	15
1.3 Προκλήσεις	16
1.4 Δομή της πτυχιακής	17
Κεφάλαιο 2 ^ο : Αλγοριθμική Λήψη Αποφάσεων	19
2.1 Σύντομη Ιστορική Αναδρομή.....	19
2.2 Αρχιτεκτονική Αλγοριθμικού Συστήματος.....	21
2.3 Λήψη Αποφάσεων – Διαδικασία	22
2.3.1 Η Ιεραρχική Ανάλυση Αποφάσεων.....	24
2.3.2 Ασαφής Λογική και Ιεραρχική Ανάλυση Αποφάσεων	27
2.4.1 Κανόνες Συσχέτισης.....	33
2.4.2 Δέντρα Αποφάσεων.....	33
2.5 Συστήματα μάθησης.....	36
2.5.1 Συστήματα Πρόσβασης Πληροφοριών	36
2.6 Δίκαιο κατά την αλγοριθμική λήψη αποφάσεων	44
Κεφάλαιο 3 ^ο : Ανθρώπινα Δικαιώματα και Διακρίσεις.....	48
3.1 Εισαγωγή.....	48
3.2 Κίνδυνοι διακρίσεων	49
3.3 Η χρήση αλγοριθμικής λήψης αποφάσεων στην Δικαιοσύνη.....	56

3.4 Ευρωπαϊκή Σύμβαση για τα Δικαιώματα του Ανθρώπου.....	59
3.5 Ευρωπαϊκό Δικαστήριο Δικαιωμάτων του Ανθρώπου	62
3.6 Εθνική Επιτροπή για τα Δικαιώματα του ανθρώπου	64
3.7 Νομοθετικό Πλαίσιο	66
3.8 Διάκριση Κατά των Γυναικών	68
Κεφάλαιο 4 ^ο : Μέτρα και εργαλεία εξάλειψης των διακρίσεων στους αλγορίθμους...	71
4.1 Εισαγωγή.....	71
4.2 Διαθέσιμα μέτρα αξιολόγησης της απόδοσης των αλγορίθμων ως προς τις διακρίσεις	71
4.2.1 Στατιστικά μέτρα	73
4.2.2 Απόλυτα μέτρα	74
4.2.3 Μέτρα υπό όρους.....	75
4.2.4 Μέτρα κατάστασης.....	76
4.3 Εργαλεία και τεχνικές αλγοριθμικής λήψης αποφάσεων με επίγνωση των διακρίσεων	76
Συμπεράσματα	83
Βιβλιογραφία	86

Κεφάλαιο 1^ο: Εισαγωγή

Στην παρούσα πτυχιακή εργασία πραγματεύεται το πεδίο της αλγοριθμικής λήψης αποφάσεων υπό το πρίσμα της πιθανής παραβίασης των ανθρωπίνων δικαιωμάτων και της πιθανής δημιουργίας διακρίσεων. Στο παρόν κεφάλαιο, πραγματοποιείται μια συνοπτική εισαγωγή στο ερευνητικό πεδίο και στο ερευνητικό θέμα. Πιο συγκεκριμένα πραγματοποιείται εισαγωγή στο πώς η αλγοριθμική λήψη αποφάσεων εφαρμόζεται στη σύγχρονη ζωή και ποια τα πιθανά οφέλη αλλά και οι ανησυχίες, στις πιθανές διακρίσεις κατά την αλγοριθμική λήψη αποφάσεων και στις προκλήσεις που απορρέουν από τη διείσδυση της αλγοριθμικής λήψης αποφάσεων στη σύγχρονη ζωή.

1.1 Αλγοριθμική λήψη αποφάσεων στη σύγχρονη κοινωνία

Σήμερα, όλο και περισσότερες αποφάσεις για τους ανθρώπους και για τους ανθρώπους λαμβάνονται χρησιμοποιώντας προγνωστικά μοντέλα που βασίζονται σε ιστορικά δεδομένα, συμπεριλαμβανομένης της βαθμολόγησης πιστοληπτικής ικανότητας, της ασφάλισης, της εξατομικευμένης τιμολόγησης και παροχής συστάσεων, του αυτοματοποιημένου ελέγχου των βιογραφικών σημειωμάτων στην εύρεση εργασίας, της αξιολόγησης του προφίλ πιθανών υπόπτων από την αστυνομία και πολλών άλλων περιπτώσεων. Η αλγοριθμική λήψη αποφάσεων βάσει δεδομένων έχει χρησιμοποιηθεί σε εφαρμογές που αφορούν φυσικά πρόσωπα για αρκετές δεκαετίες. Για παράδειγμα, οι αλγόριθμοι βαθμολόγησης της πιστοληπτικής ικανότητας αναπτύσσονταν στην πράξη ήδη από τη δεκαετία του 50' (Furletti, 2002). Οι δε αλγόριθμοι εκτίμησης κινδύνου στα δικαστικά συστήματα είναι πιο πρόσφατοι και χρησιμοποιούνται από τη δεκαετία του 70' (Kehl, et al., 2017).

Ωστόσο, με την έλευση ολοένα πιο εξελιγμένων και πολύπλοκων μεθόδων μηχανικής μάθησης, αλγορίθμων, εξόρυξης δεδομένων και άλλων τεχνικών, και με τη διεύρυνση της δημιουργίας και διαθεσιμότητας «μεγάλων δεδομένων», η αλγοριθμική λήψη αποφάσεων διεισδύει ολοένα περισσότερο σε έναν συνεχώς αυξανόμενο αριθμό ανθρωποκεντρικών εφαρμογών. Στις εφαρμογές αυτές όπου μέχρι πρότινος οι αποφάσεις λαμβάνονταν αποκλειστικά από ανθρώπους, οι

αλγόριθμοι χρησιμοποιούνται για να βοηθήσουν ή μερικές φορές ακόμη και να υποκαταστήσουν πλήρως τους ανθρώπινους. Ως εκ τούτου, ως αλγοριθμική λήψη αποφάσεων μπορεί να αναφέρεται η υποβοηθούμενη από αλγορίθμους λήψη απόφασης από ανθρώπους και η αμιγώς οδηγούμενη από αλγορίθμους λήψη απόφασης.

Οι Köchling και Wehner (2020) αναφέρουν μεταξύ άλλων την εφαρμογή της αλγοριθμικής λήψης αποφάσεων στο πεδίο της Διοίκησης Ανθρώπινου Δυναμικού, με παράδειγμα την περίπτωση όπου η αρχική αξιολόγηση των βιογραφικών σημειωμάτων υποψηφίων για θέσεις εργασίας πραγματοποιείται με εφαρμογή εξελιγμένων αλγορίθμων (Köchling & Wehner, 2020). Οι Alikhademi et al. (2022) αναφέρουν τη διείσδυση της αλγοριθμικής λήψης αποφάσεων στο πεδίο της ασφάλειας με το ρηξικέλευθο παράδειγμα της προβλεπτικής αστυνόμευσης, όπου δημιουργείται και παρακολουθείται το προφίλ υπόπτων για τέλεση αδικημάτων (Alikhademi, et al., 2022). Οι Covington et al. (2016) εξέτασαν τη διείσδυση της αλγοριθμικής λήψης αποφάσεων για την εξατομίκευση των προϊόντων και των διαφημίσεων, ως επί το πλείστο στα μέσα κοινωνικής δικτύωσης και στα κοινωνικά μέσα, προτείνοντας ότι τα πλεονεκτήματα που κομίζει η αλγοριθμική λήψη αποφάσεων υπερκερνούν τα μειονεκτήματα και τις ανησυχίες περί βλάβης της ιδιωτικότητας (Covington, et al., 2016). Οι δε Brown et al. (2019) εξετάζουν ενδελεχώς τη διείσδυση και εφαρμογή της αλγοριθμικής λήψης αποφάσεων στο δημόσιο τομέα, με έμφαση στο κοινωνικό κράτος (Brown, et al., 2019).

Η διείσδυση των τεχνολογιών μηχανικής μάθησης, των τεχνολογιών της επιστήμης των δεδομένων, συμπεριλαμβανομένης της εξόρυξης δεδομένων, έχει καταφανώς οδηγήσει σε αύξηση των αποφάσεων οι οποίες λαμβάνονται καθημερινά από αλγορίθμους ή με τη βοήθεια αλγορίθμων (Bader & Kaiser, 2019). Αυτό έχει με τη σειρά του ευαισθητοποιήσει το κοινό, εγείροντας ανησυχίες για πιθανή παραβίαση των ανθρωπίνων δικαιωμάτων και δημιουργία διακρίσεων εις βάρος συγκεκριμένων ομάδων ανθρώπων (Žliobaitė, 2017a). Μια τέτοιου είδους διάκριση, ειδικά υπό το πρίσμα της αυτοματοποιημένης λήψης της απόφασης, είναι κατά πάσα πιθανότητα όχι μόνον ακούσια αλλά και απρόσμενη, λαμβάνοντας υπόψη ως δεδομένο ότι οι

αξιοποιούμενοι αλγόριθμοι είναι εγγενώς αντικειμενικοί (Lepri, et al., 2018). Ωστόσο, το πλέον ανησυχητικό είναι το γεγονός ότι η αλγοριθμική λήψη αποφάσεων, παρά το γεγονός ότι πραγματοποιείται στη βάση ιστορικών δεδομένων, μπορεί να οδηγήσει σε παραβίαση των ανθρωπίνων δικαιωμάτων και δημιουργία διακρίσεων ακόμη και αν οι αξιοποιούμενοι αλγόριθμοι είναι εγγενώς αντικειμενικοί, και η όλη διαδικασία είναι δίκαιη και καλοπροαίρετη (De Laat, 2018). Αυτό οφείλεται στο γεγονός ότι τα ιστορικά δεδομένα τα οποία λαμβάνονται υπόψη προκειμένου να τρέξει ένας προβλεπτικός αλγόριθμος δεν μπορεί να είναι πάντα δίκαια και αντικειμενικά, αλλά και επειδή αντικατοπτρίζουν το μέσο πληθυσμό αφήνουν μεγάλο χώρο για αποκλίσεις κατά περίπτωση διαφορετικών ανθρώπων (Calders & Zliobaite, 2013).

Απόρροια του παραπάνω τεκμηριωμένου προβληματισμού ήταν η ανάδειξη ενός νέου κλάδου της αλγοριθμικής λήψης αποφάσεων ή και μηχανικές μάθησης ή και εξόρυξης δεδομένων με επίγνωση των διακρίσεων, όπου οι ακαδημαϊκοί εστιάζουν σε τρόπους αποτροπής πιθανών διακρίσεων οι οποίες μπορεί να δημιουργηθούν από την εφαρμογή των επιλεγμένων αλγορίθμων (Xu, et al., 2020). Προκειμένου για την ανάπτυξη αυτού του κλάδου, απαιτείται ο αυστηρός ορισμός του τι είναι διάκριση, ποιο είναι το σύνολο των ανθρωπίνων δικαιωμάτων τα οποία πρέπει να γίνονται σεβαστά κατά την εφαρμογή της αλγοριθμικής λήψης αποφάσεων, και κατόπιν τα παραπάνω πρέπει να εισάγονται υπό μορφή κανόνων στο σώμα κάθε αλγόριθμου (Pedreshi, et al., 2008). Μάλιστα, λαμβάνοντας υπόψη το γεγονός ότι διαφορετικές ομάδες του πληθυσμού μπορεί να υπόκεινται σε διαφορετικής φύσεως διακρίσεις, η παραπάνω ανάλυση πρέπει να γίνεται σε επίπεδο αρχικά πληθυσμού, και έπειτα σε επίπεδο ομάδων κοινών δημογραφικών χαρακτηριστικών. Μολονότι παραπάνω αναφέρεται πως πρέπει να οριστεί τι είναι διάκριση και τι είναι παραβίαση των ανθρωπίνων δικαιωμάτων, εντούτοις οι ορισμοί αυτοί απορρέουν από τη διεθνή και εθνική νομοθεσία, με αποτέλεσμα το αντικείμενο της αλγοριθμικής λήψης αποφάσεων με επίγνωση των διακρίσεων να μετατρέπεται σε ένα πρόβλημα μετάφρασης των κανονισμών αυτών προκειμένου να είναι εφικτή η ενσωμάτωσή τους στους αλγορίθμους (Corbett-Davies, et al., 2017). Επιπρόσθετα, στα πλαίσια της

αλγοριθμικής λήψης αποφάσεων με επίγνωση των διακρίσεων, έντονο ενδιαφέρον συγκεντρώνει το πεδίο του εντοπισμού και της εξάλειψης όχι των ίδιων των διακρίσεων και των ίδιων των παραβιάσεων των ανθρωπίνων δικαιωμάτων, αλλά των αιτίων που δημιουργούν τις διακρίσεις ή παραβιάσεις προκειμένου να καθίσταται εφικτή η διαρκής βελτίωση της αντικειμενικότητας των εφαρμοζόμενων αλγορίθμων και τεχνικών μηχανικής μάθησης. Ένας σημαντικός κίνδυνος κατά την αντιμετώπιση της ανησυχίας του κοινού είναι η λήψη τόσων περιοριστικών κανονισμών οι οποίοι θα οδηγήσουν τόσο εφαρμοζόμενους ο αριθμός και τις εφαρμοζόμενες τεχνικές μηχανικής μάθησης ή τεχνητής νοημοσύνης σε τέλμα μην επιτρέποντας την αποδοτική αλγοριθμική λήψη αποφάσεων. στα πλαίσια αυτά εντοπίζεται έντονο ενδιαφέρον γύρω από την ανάπτυξη μέτρων για την ανίχνευση διακρίσεων και παραβιάσεων των ανθρωπίνων δικαιωμάτων και γύρω από την ανάπτυξη μέτρων της απόδοσης των εφαρμοζόμενων αλγορίθμων και τεχνικών τεχνητής νοημοσύνης σε ότι αφορά στην αποφυγή των διακρίσεων (Starke, et al., 2022). Σε ό,τι αφορά στην εφαρμογή της αλγοριθμικής λήψης αποφάσεων στην ιατρική και δη στην προσυμπτωματική ή προληπτική ιατρική καταγράφονται σημαντικά οφέλη σε όρους βελτίωσης της ακρίβειας της διάγνωσης αλλά και σε όρους βελτίωσης της ταχύτητας της διάγνωσης, ειδικά σε περιπτώσεις όπου μέσω της διαφορικής διάγνωσης η αντιμετώπιση του ασθενούς θα καθυστερούσε σημαντικά. Ωστόσο, σύμφωνα με τους Grote και Berens (2020), εγείρονται σημαντικά ηθικά ζητήματα σε ότι αφορά στην εφαρμογή της αλγοριθμικής λήψης αποφάσεων στην ιατρική, συμπεριλαμβανομένων και ζητημάτων πιθανής δημιουργίας διάκρισης. Η διάκριση στο πεδίο αυτό μπορεί να είναι είτε υπό μορφή αποκλεισμού μερίδας των ασθενών από καινοτόμες μεθόδους αλγοριθμικά υποβοηθούμενης διάγνωσης είτε υπό μορφή άδικης μεταχείρισης των ασθενών από τα συστήματα υγείας στη βάση της ομαδοποίησης τους βάση τις αρχικής διάγνωσης (Grote & Berens, 2020).

Η αλγοριθμική λήψη αποφάσεων είναι πολλά υποσχόμενη σε ότι αφορά στη βελτίωση της ακρίβειας αλλά και της επεκτασιμότητας και προσαρμοστικότητας των υφιστάμενων εφαρμογών της. Αυτό της προσδίδει ιδιαίτερη σημασία όταν πρόκειται να εφαρμοστεί σε τομείς όπου η ακρίβεια, η προσαρμοστικότητα και η ταχύτητα είναι

μείζονος σημασίας, όπως είναι η ιατρική. Αναφέρεται ότι η αλγοριθμική λήψη αποφάσεων και δη η εφαρμογή της τεχνητής νοημοσύνης για την αλγοριθμική λήψη αποφάσεων οδηγεί σε βελτίωση της ταχύτητας και της ακρίβειας της διάγνωσης ασθενών με καρκίνο. Αξίζει να αναφερθεί ότι σύμφωνα με μελέτη των Duggento et al. (2021), βελτιώθηκε η ακρίβεια και η ταχύτητα διάγνωσης και σταδιοποίησης καρκίνων του μαστού με εφαρμογή τεχνικών τεχνητής νοημοσύνης και συγκεκριμένα βαθιάς μάθησης, ενώ σύμφωνα με μελέτη των Djuric et al. (2017) η εφαρμογή βαθιάς μάθησης μπορεί να οδηγήσει σε βελτίωση της αποτελεσματικότητας και της αποδοτικότητας της ιστοπαθολογίας καθιστώντας πιο εφικτή και πιο γρήγορη την παροχή εξατομικευμένης θεραπείας σε καρκινοπαθείς (Duggento, et al., 2021; Djuric, et al., 2017).

Σύμφωνα με τον Bell (2021) η αξιοποίηση της αλγοριθμικής λήψης αποφάσεων μπορεί να οδηγήσει σε καλύτερη ακρίβεια εντοπισμού παραβατών, π.χ. πολιτών που οπλοφορούν, οδηγώντας τόσο σε μείωση των άσκοπων προληπτικών ελέγχων όσο και σε βελτίωση σε όρους μείωσης της εγκληματικότητας (Bell, 2021). Οι Levy et al. (2021) προτείνουν ότι μπορεί να μειωθεί η αποτελεσματικότητα της πρόληψης της εγκληματικότητας και πρόβλεψης άλλων κοινωνικών φαινομένων με εφαρμογή της αλγοριθμικής λήψης αποφάσεων (Levy, et al., 2021). Οι Langer και Lander (2021) προτείνουν ότι η εφαρμογή της αλγοριθμικής λήψης αποφάσεων στη διοίκηση ανθρώπινου δυναμικού μπορεί να οδηγήσει σε σημαντική μείωση των οικειοθελών αποχωρήσεων, απόρροια της καλύτερης αντιστοίχισης υποψηφίων σε ανοιχτές θέσεις εργασίας (Langer & Landers, 2021). Σε ό,τι αφορά στο πεδίο της εκπαίδευσης, ο Prinsloo (2020) προτείνει ότι η αξιοποίηση της αλγοριθμικής λήψης αποφάσεων μπορεί να οδηγήσει σε βελτίωση του σχεδιασμού των προγραμμάτων εκπαίδευσης και εν γένει σε βελτίωση της εμπειρίας εκπαίδευσης που προσφέρεται στους μαθητές και φοιτητές (Prinsloo, 2020).

Σύμφωνα με τους οι Dobbie et al. (2018), η εφαρμογή της αλγοριθμικής λήψης αποφάσεων μπορεί να οδηγήσει επί τούτου σε μείωση της προκατάληψης στο δικαστικό σύστημα, καθώς θα αποκλειστούν τυχόν έμφυτες προκαταλήψεις των ίδιων των δικαστικών (Dobbie, et al., 2018)

1.2 Διακρίσεις στην αλγοριθμική λήψη αποφάσεων

Όπως προαναφέρεται, παρά τα πολλά και σημαντικά πλεονεκτήματα της αλγοριθμικής λήψης αποφάσεων εγείρονται σημαντικές ανησυχίες αναφορικά με πιθανές διακρίσεις οι οποίες μπορεί να γίνονται σε βάρος ατόμων και ομάδων του πληθυσμού με ορισμένα κοινά κοινωνικά πολιτιστικά ή άλλα χαρακτηριστικά τα οποία διαφοροποιούν αυτή τη μερίδα του πληθυσμού από τον υπόλοιπο πληθυσμό. Όπως προαναφέρεται στην προηγούμενη παράγραφο, δεδομένου ότι οι αλγόριθμοι για τη λήψη αποφάσεων σχεδιάζονται με βάση την εκτέλεση προβλεπτικών μοντέλων αξιοποιώντας ιστορικά δεδομένα, εισάγεται κίνδυνος δημιουργίας ανισοτήτων και διακρίσεων δεδομένου ότι τα ιστορικά δεδομένα αφορούν και χαρακτηρίζουν το μέσο πληθυσμό και δε γίνεται άμεση διάκριση για διαφορετικές και ειδικές ομάδες του πληθυσμού. χαρακτηριστικά αναφέρεται ότι οι διακρίσεις που μπορεί να λάβουν χώρα μέσω της αλγοριθμικής λήψης αποφάσεων μπορεί να γίνουν στη βάση του διαφορετικού φύλου, καταγωγής, θρησκείας, εκπαιδευτικού επιπέδου, φυλής κ.α..

Ο Sweeney (2013) εντοπίζει μεροληψία στην πλατφόρμα διαφημίσεων της Google, όπου ονόματα αφροαμερικανών συσχετίζονται με αρχεία συλλήψεων συγκριτικά με ονόματα λευκών αμερικανών, ενώ οι Bolukbasi et al. (2016) εντοπίζουν την εγγενή μεταφορά στερεοτύπων στους αλγορίθμους λήψης αποφάσεων απλά και μόνον με τη χρήση λέξεων οι οποίες χαρακτηρίζουν το φύλο (Sweeney, 2013; Bolukbasi, et al., 2016).

Σύμφωνα με τον Griffard (2019) το εργαλείο COMPAS το οποίο εφαρμόζεται σε διάφορες πολιτείες των ΗΠΑ για την αξιολόγηση της πιθανότητας υποτροπής υπόδικων ή κατάδικων είναι έντονα προκατειλημμένο εναντίον αφροαμερικανών και προτείνει ότι το εργαλείο Patternizr που χρησιμοποιείται από τις αρχές στη Ν. Υόρκη έχει απεμπολήσει αυτές τις προκαταλήψεις (Griffard, 2019). Ωστόσο, σύμφωνα με άρθρο του Kirkpatrick (2017) επιβεβαιώνονται τα αναφερόμενα στην προηγούμενη παράγραφο, καθώς προτείνεται ότι η μεροληψία του εργαλείου COMPAS οφείλεται στην μεροληψία των δεδομένων που έχουν εισαχθεί και εισάγονται στο σύστημα και όχι στον ίδιο τον εφαρμοζόμενο αλγόριθμο (Kirkpatrick, 2017).

1.3 Προκλήσεις

Η συγγραφή της πτυχιακής μου σε ένα θέμα τόσο περίπλοκο όπως η αλγοριθμική λήψη αποφάσεων, οι διακρίσεις και τα ανθρώπινα δικαιώματα ήταν ένα δύσκολο έργο. Σε όλη αυτή τη διαδικασία, έχω αντιμετωπίσει ποικίλες προκλήσεις, που κυμαίνονται από την σύλληψη του ερευνητικού ερωτήματος, τη συλλογή και ανάλυση δεδομένων, έως την παραγωγή ενός καλοδουλεμένου γραπτού προϊόντος. Η ακόλουθη ενότητα θα συζητήσει τις συγκεκριμένες προκλήσεις που αντιμετώπισα κατά τη συγγραφή αυτής της πτυχιακής.

1. Εννοιολόγηση του ερωτήματος της έρευνας: Η αλγοριθμική λήψη αποφάσεων και οι επιπτώσεις της στα ανθρώπινα δικαιώματα είναι ένα περίπλοκο θέμα που απαιτεί προσεκτική σκέψη και εξέταση. Ως εκ τούτου, χρειάστηκε να εξετάσω προσεκτικά το ερευνητικό ερώτημα που θα αντιμετωπιστεί σε αυτήν την εργασία. Αντιμέτωπα αρκετές προκλήσεις κατά τη σύλληψη του ερευνητικού ερωτήματος. Πρώτα, έπρεπε να αποφασίσω σε ποιες συγκεκριμένες πτυχές της αλγοριθμικής λήψης αποφάσεων και στις επιπτώσεις της στα ανθρώπινα δικαιώματα να επικεντρωθώ. Δεύτερον, έπρεπε να προσδιορίσω την καλύτερη προσέγγιση για την ανάλυση του θέματος, όπως μια ποιοτική ή ποσοτική προσέγγιση. Τέλος, χρειάστηκε να σκεφτώ πώς να δομήσω αποτελεσματικά την εργασία και να παρουσιάσω τα ευρήματα.
2. Συγκέντρωση και ανάλυση δεδομένων: Μια δεύτερη σειρά προκλήσεων προέκυψε όταν άρχισα να συνεχίζω την έρευνά μου και να συλλέγω δεδομένα για την εργασία. Έπρεπε να εξετάσω προσεκτικά ποιες πηγές δεδομένων να χρησιμοποιήσω και πώς να αναλύσω αποτελεσματικά τα δεδομένα για να υποστηρίξω το ερώτημα της έρευνας. Αυτή η διαδικασία περιλάμβανε την εξέταση μιας ποικιλίας ηθικών κριτηρίων, καθώς αντιμετώπισα την πρόκληση της προστασίας των δικαιωμάτων εκείνων των οποίων τα δεδομένα χρησιμοποιήθηκαν. Επιπλέον, έπρεπε να εξετάσω την πιθανή μεροληψία στα δεδομένα που συλλέγονται και να αναλύσω τα δεδομένα με ανοιχτό μυαλό.

3. Δημιουργία του γραπτού προϊόντος: Αφού συγκεντρώθηκαν και αναλύθηκαν τα δεδομένα, έπρεπε να σκεφτώ πώς να παρουσιάσω τα ευρήματα σε ένα καλοφτιαγμένο γραπτό προϊόν. Αυτή η διαδικασία απαιτήσε να εξετάσω προσεκτικά πώς να δομήσω αποτελεσματικά το έγγραφο και να παρουσιάσω τα ευρήματα με σαφή, συνοπτικό και ελκυστικό τρόπο. Επιπλέον, έπρεπε να σκεφτώ πώς να αναφέρω σωστά τις πηγές που χρησιμοποιήθηκαν στην εργασία, καθώς και πώς να επικοινωνήσω αποτελεσματικά τις επιπτώσεις των ευρημάτων της εργασίας.

Συνολικά, η συγγραφή μιας διατριβής για ένα θέμα τόσο περίπλοκο όπως η αλγοριθμική λήψη αποφάσεων, οι διακρίσεις και τα ανθρώπινα δικαιώματα ήταν μια πρόκληση. Αντιμέτωπισα ποικίλες προκλήσεις κατά τη σύλληψη του ερευνητικού ερωτήματος, τη συλλογή και ανάλυση δεδομένων και τη δημιουργία του γραπτού προϊόντος.

1.4 Δομή της πτυχιακής

Η παρούσα πτυχιακή εργασία αποτελείται από πέντε κεφάλαια. Στο πρώτο, παρόν κεφάλαιο, πραγματοποιείται μια συνοπτική εισαγωγή στο ερευνητικό πεδίο και στο ερευνητικό θέμα. Πιο συγκεκριμένα πραγματοποιείται εισαγωγή στο πώς η αλγοριθμική λήψη αποφάσεων εφαρμόζεται στη σύγχρονη ζωή και ποια τα πιθανά οφέλη αλλά και οι ανησυχίες, στις πιθανές διακρίσεις κατά την αλγοριθμική λήψη αποφάσεων και στις προκλήσεις που απορρέουν από τη διείδυση της αλγοριθμικής λήψης αποφάσεων στη σύγχρονη ζωή. Στο δεύτερο κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από τη βιβλιογραφική ανασκόπηση στο πεδίο της αλγοριθμικής λήψης αποφάσεων, όπου μετά από μια σύντομη ιστορική αναδρομή περιγράφεται η αρχιτεκτονική ενός αλγοριθμικού συστήματος λήψης αποφάσεων, περιγράφεται η όλη διαδικασία λήψης αποφάσεων. Στη συνέχεια γίνεται αναφορά στα συστήματα μάθησης, με έμφαση στην εφαρμογή τους με διαφορετικές πηγές δεδομένων. Στο τρίτο κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από την ανασκόπηση στο πεδίο των ανθρωπίνων δικαιωμάτων και των διακρίσεων, υπό το πρίσμα της αλγοριθμικής λήψης αποφάσεων. Περισσότερη έμφαση δίδεται στην εφαρμογή της αλγοριθμικής λήψης αποφάσεων στη δικαιοσύνη, με εκτενή ανάλυση

του συνόλου των συμβάσεων και νομοθετικών πλαισίων τα οποία διέπουν την απονομή δικαιοσύνης και τα σχετικά ανθρώπινα δικαιώματα. Στο τέταρτο κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από τη βιβλιογραφική ανασκόπηση στο πεδίο των διακρίσεων και του σεβασμού των ανθρωπίνων δικαιωμάτων κατά την εφαρμογή της αλγοριθμικής λήψης αποφάσεων στη δικαιοσύνη, τόσο υπό το πρίσμα των προτεινόμενων μέτρων για την εξάλειψη των διακρίσεων όσο και υπό το πρίσμα των διαθέσιμων μέσων για τη μέτρηση της απόδοσης των αλγορίθμων σε ό, τι αφορά στην αποφυγή διακρίσεων. Τέλος στο πέμπτο κεφάλαιο εξάγονται συμπεράσματα κατόπιν ολοκλήρωσης της βιβλιογραφικής ανασκόπησης, συζητούνται περιορισμοί των συμπερασμάτων αυτών και δίνονται προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 2^ο: Αλγοριθμική Λήψη Αποφάσεων

Σε αυτό το κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από τη βιβλιογραφική ανασκόπηση στο πεδίο της αλγοριθμικής λήψης αποφάσεων, όπου μετά από μια σύντομη ιστορική αναδρομή περιγράφεται η αρχιτεκτονική ενός αλγοριθμικού συστήματος λήψης αποφάσεων, περιγράφεται η όλη διαδικασία λήψης αποφάσεων. Στη συνέχεια γίνεται αναφορά στα συστήματα μάθησης, με έμφαση στην εφαρμογή τους με διαφορετικές πηγές δεδομένων.

2.1 Σύντομη Ιστορική Αναδρομή

Η λήψη αποφάσεων ξεκινά με την εμφάνιση του ανθρώπου, ενώ έλλογα απαντάται μόνο στο ανθρώπινο είδος, καθώς στα υπόλοιπα μέλη του ζωικού βασιλείου πραγματοποιείται ενστικτωδώς και απορρέει βασικά από το ένστικτο της επιβίωσης. Η λήψη αποφάσεων συνιστά μια πολυδιάστατη διαδικασία ενώ πραγματοποιείται τόσο συνειδητά όσο και ασυνείδητα καθώς λαμβάνεται υπόψιν το σύνολο των παραγόντων που επηρεάζουν την υπό εξέταση υπόθεση. Η λήψη της απόφασης επηρεάζεται από την φύση και το χαρακτήρα του υποκειμένου αλλά και από την φύση και την σοβαρότητα του υπό εξέταση ζητήματος, ενώ με μία πρώτη ματιά δεν είναι ορατό ότι αποτελεί μια σύνθετη μαθηματική διαδικασία.

Ως πρώτη προσπάθεια επιστημονικής αντιμετώπισης του ζητήματος δύναται να θεωρηθεί η καταγραφή του Paerto (1896), ο οποίος θέτει θεμελιώδεις αξιωματικές βάσεις ενώ για πρώτη φορά εισάγει και την έννοια της αποτελεσματικότητας όσον αφορά την ληφθήσα απόφαση σε συνάρτηση με το υπό εξέταση ζήτημα. Αρκετά χρόνια αργότερα, ο Koopmans (1951) επεκτείνει την έννοια της αποτελεσματικότητας με την εισαγωγή του όρου «αποτελεσματικό σύνολο» [1]. Έτσι, σήμερα ένας κοινά αποδεκτός ορισμός όσον αφορά τη λήψη αποφάσεων είναι ο ακόλουθος:

«Ο όρος λήψη αποφάσεων αναφέρεται στην διαδικασία επιλογής διαφορετικών εναλλακτικών λύσεων και τρόπων δράσης με στόχο την επίτευξη ενός ή περισσότερων στόχων»[2].

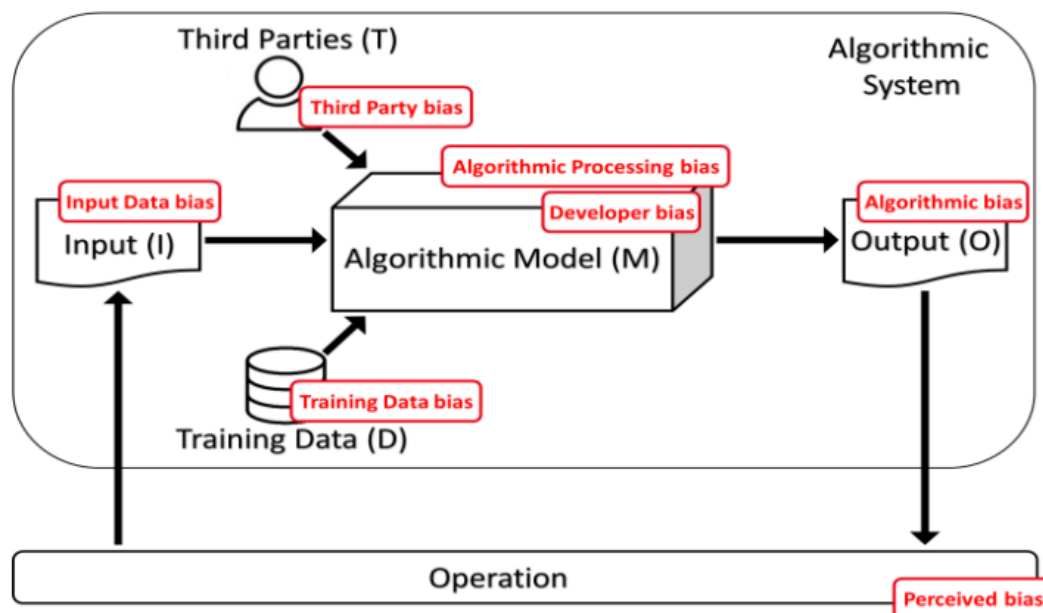
Ταυτόχρονα, η λήψη αποφάσεων αποτελεί έναν επιστημονικό κλάδο ιδιαίτερα διευρυμένο και διαρκώς αναπτυσσόμενο καθώς η λήψη αποφάσεων εισήχθη στην τεχνολογία και στις μηχανές που χρησιμοποιούμε προς διευκόλυνση της καθημερινότητάς μας, οι οποίες θεμελιώνονται από αλγόριθμους τεχνητής νοημοσύνης. Οι αλγόριθμοι που αναπτύσσονται κωδικοποιούν την διαδικασία λήψης της απόφασης μέσα από την δημιουργία ενός μηχανισμού πρόβλεψης και επιλογής των δυνατών αποτελεσμάτων. Μέρος της διαδικασίας αποτελεί η αξιολόγηση τόσο των βραχυπρόθεσμων όσο και των μακροπρόθεσμων αποτελεσμάτων [1].

Στελέχη επιχειρήσεων, διατείνονται ότι η τέχνη λήψης αποφάσεων αποτελεί ένα επίκτητο χαρακτηριστικό το οποίο αναπτύσσεται με την πάροδο των ετών και στηρίζεται στην εμπειρία που ο άνθρωπος αποκτά τόσο από τις ορθές αποφάσεις, όσο και από την αξιολόγηση των λανθασμένων επιλογών (Learning by trial and error). Για αρκετά χρόνια, οι επιχειρηματικές αποφάσεις στηριζόντουσαν κατά αποκλειστικότητα στα λογιστικά δεδομένα και στις λογιστικές αναλύσεις, οπότε ήταν κοινά αποδεκτό ότι η λογιστική είναι ο μοναδικός επιστημονικός κλάδος που συνδέεται με την λήψη αποφάσεων.

Εν τούτοις, το πέρασμα των ετών και η εξέλιξη διαφόρων επιστημονικών χώρων ανέδειξε την συμβολή και άλλων επιστημονικών κλάδων στη λήψη αποφάσεων, καθώς αυτή στηρίζεται από μια πληθώρα μεθοδολογικών προσεγγίσεων. Έτσι, πλέον είναι αποδεκτό ότι η λήψη αποφάσεων συνδυάζει θεωρίες και στρατηγικές μεθόδους από τον χώρο της Διοικητικής Επιστήμης, της Ψυχολογίας και των Τεχνολογιών της Πληροφορικής. Ιδιαίτερα με την ευρεία χρήση πληροφοριακών συστημάτων δίνεται η δυνατότητα επιλογής πολλών και διαφορετικών λύσεων σε πλήθος προβλημάτων, ενώ ολοένα και αποτελεσματικότερος σχεδιασμός αλγοριθμικών συστημάτων λήψης αποφάσεων, αποτελεί μια επιταγή των σύγχρονων οικονομικοκοινωνικών συστημάτων καθώς η όποια λανθασμένη επιλογή κοστίζει [3].

2.2 Αρχιτεκτονική Αλγοριθμικού Συστήματος

Στην Εικόνα 1, παρουσιάζεται η πιο απλή δομή που αφορά ένα αλγοριθμικό σύστημα το οποίο χρησιμοποιεί τεχνικές μάθησης μηχανών (machine learning techniques).



Εικόνα 1: Αλγοριθμικό σύστημα το οποίο χρησιμοποιεί τεχνικές μάθησης μηχανών

- **Είσοδος (Input Data bias (I)):** πρόκειται για την εισαγωγή δεδομένων από τον χρήστη με σκοπό την εκτέλεση μιας συγκεκριμένης εργασίας από το σύστημα. Παράδειγμα αποτελεί η αναζήτηση που πραγματοποιούν οι χρήστες του διαδικτύου στις μηχανές αναζήτησης. Σε κάθε περίπτωση, ο χρήστης εισάγει συγκεκριμένες λέξεις ενώ το σύστημα του εμφανίζει όλες τις σχετικές με το ζήτημα πληροφορίες.
- **Δεδομένα εκμάθησης (Training Data (D)):** χρησιμότητα των δεδομένων εκμάθησης είναι η ενημέρωση του αλγοριθμικού μοντέλου με στοιχεία που θα βοηθήσουν στις διεργασίες εκτέλεσης του αλγορίθμου, με αντίστοιχη κατάταξη των δεδομένων εκμάθησης. Για παράδειγμα, κατά την αναζήτηση πληροφοριών στο διαδίκτυο, τροφοδοτείται με δεδομένα εκμάθησης και ταυτόχρονα αναπτύσσεται ο αλγόριθμος κατάταξης για την παρουσίαση πληροφοριών, τύπων ιστοσελίδων που αναζήτησε ο χρήστης.

- **Τρίτοι (Third Parties (T)):** Ως τρίτο μέρος, νοείται οποιοσδήποτε άλλος εκτός από τους άμεσα εμπλεκόμενους σε ένα σύστημα. Κυρίως, είναι ρυθμιστικές αρχές ή άλλοι παράγοντες που έχουν ενεργό ρόλο ενώ επηρεάζουν την χρήση και τα αποτελέσματα του συστήματος. Χαρακτηριστικό παράδειγμα είναι οι μηχανές αναζήτησης στο διαδίκτυο όπου περιορίζουν τα αποτελέσματα που εμφανίζονται στους χρήστες ώστε να εναρμονίζονται με τους κανόνες που ισχύουν σε κάθε χώρα σχετικά με την διασφάλιση και την προστασία των προσωπικών δεδομένων.
- **Αλγόριθμος (Algorithm (M)):** Πρόκειται για το «μυαλό» του αλγοριθμικού συστήματος, όπου λειτουργεί με βάση ένα αιτιοκρατικό και δυναμικό μοντέλο εκτελώντας μια αλληλουχία αλγορίθμων με αντίστοιχες ενέργειες και διεργασίες, συνδυάζοντας την είσοδο, τα δεδομένα εκμάθησης και τα τρίτα μέρη, με σκοπό την δημιουργία αξιόπιστου αποτελέσματος, δηλαδή την έξοδο. Στα παραδείγματα που έχουν αναφερθεί σχετικά με τις μηχανές αναζήτησης στο διαδίκτυο ο πιο διαδεδομένος αλγόριθμος που χρησιμοποιείται είναι ο αλγόριθμος κατάταξης, με βάση τον οποίο αποδίδεται μια βαθμολογία για κάθε ιστοσελίδα που προβάλεται στον χρήστη ανάλογα με τις λέξεις κλειδιά τις οποίες έχει χρησιμοποιήσει.
- **Έξοδος (Output (O)):** Η έξοδος του αλγοριθμικού συστήματος είναι το τελικό αποτέλεσμα το οποίο προκύπτει από τις προσλαμβάνουσες τιμές του χρήστη. Όπως για παράδειγμα η ιεράρχιση των αποτελεσμάτων που λαμβάνει ο χρήστης ύστερα από μια αναζήτηση στο διαδίκτυο.

2.3 Λήψη Αποφάσεων – Διαδικασία

Η διαδικασία σχεδιασμού ενός συστήματος λήψης αποφάσεων περιλαμβάνει ένα σύνολο βημάτων. Προκειμένου να προσδιορίσουμε τα βήματα αυτά κρίνεται ορθό να γίνει αναφορά στην έννοια της *απόφασης*. Απόφαση καλείται η εκτέλεση μιας πράξης που απορρέει από σκέψη και λαμβάνεται κατόπιν επεξεργασίας όλων των δυνατών επιλογών. Σε περιπτώσεις που οι δυνατές επιλογές είναι περισσότερες από μια, η λήψη της απόφασης συνίσταται στην απόρριψη όλων των δυνατών επιλογών εκτός μίας. Η απόφαση αποτελεί νέα γνώση, καθώς πρόκειται για αποτέλεσμα σκέψης και επεξεργασίας δεδομένων [1].

Εν τούτοις, μια απόφαση που φαίνεται ορθή δεν οδηγεί πάντα στο ανάλογο αποτέλεσμα, καθώς η αποτελεσματικότητά της επηρεάζεται από ένα σύνολο παραγόντων όπως ο χρόνος λήψης της, η ιδιοσυγκρασία και ο χαρακτήρας του ανθρώπου που θα την υλοποιήσει αλλά και την μορφή πρακτικής που θα έχει. Τα βήματα αυτά συνιστούν τέσσερις φάσεις[4].

Η πρώτη φάση περιλαμβάνει την διατύπωση του προβλήματος με σαφήνεια καθώς και την καταγραφή όλων των πιθανών εναλλακτικών λύσεων **Intelligence Phase**. Επίσης, στη φάση αυτή λαμβάνονται υπόψιν οι συνθήκες λήψης της απόφασης καθώς και η δυνατότητα συλλογής επαρκούς αριθμού δεδομένων. Γίνεται κατανοητό, ότι ο μη ορθός προσδιορισμός του προβλήματος θα οδηγήσει σε εσφαλμένα αποτελέσματα. Ο καθορισμός του προβλήματος ακολουθείται από την αναγνώριση των στόχων. Σε αρκετές περιπτώσεις, όπως για παράδειγμα στον επιχειρηματικό κόσμο, οι στόχοι είναι σαφείς, εν τούτοις δεν συμβαίνει το ίδιο σε όλες τις περιπτώσεις ενώ σημαντικό είναι οι στόχοι να βρίσκονται σε συνάρτηση με τις αντίστοιχες διαθέσιμες λύσεις [4].

Η δεύτερη φάση περιλαμβάνει την καταγραφή και αναγνώριση όλων των εναλλακτικών λύσεων, **Design Phase**. Καθίσταται σαφές ότι το πλήθος των εναλλακτικών λύσεων είναι ευθέως ανάλογο με την δυσκολία εύρεσης της κατάλληλης λύσης. Από την άλλη πλευρά, η ύπαρξη πλήθους λύσεων δεν συνεπάγεται στην ανάπτυξη ενός αξιόπιστου συστήματος λήψης αποφάσεων[4].

Στο τρίτο στάδιο πραγματοποιείται η αξιολόγηση των διαθέσιμων εναλλακτικών λύσεων, **Choice Phase**, καθώς είναι πολύ πιθανό λύσεις που αρχικά δείχνουν ορθές, με μια προσεκτικότερη μελέτη να μην είναι. Η αξιολόγηση των διαθέσιμων λύσεων είναι δυνατόν να πραγματοποιηθεί μέσω εμπειρίας, μέσω πειραματισμού ή μέσω έρευνας. Από τους παραπάνω τρόπους αξιολόγησης, ο πιο άμεσος και ευρέως χρησιμοποιούμενος είναι η αξιολόγηση μέσω εμπειρίας, τρόπος στον οποίο εισάγεται το υποκειμενικό κριτήριο [4].

Στην περίπτωση ανάπτυξης αλγοριθμικών συστημάτων, η επιλογή της καταλληλότερης λύσης πραγματοποιείται σε συνάρτηση με το σύνολο των δεδομένων που εισήχθησαν. Έτσι, το μοντέλο λαμβάνει υπόψιν κριτήρια και περιορισμούς που εισήχθησαν σε προηγούμενη φάση.

Στο τελικό στάδιο, **Implementation Phase**, πραγματοποιείται η υλοποίηση της απόφασης [4].

2.3.1 Η Ιεραρχική Ανάλυση Αποφάσεων

Η επίλυση προβλημάτων και η λήψη αποφάσεων αποτελούν διαδικασίες οι οποίες αντιμετωπίζονται από πλήθος επαγγελματικών κλάδων, όπως οι μηχανικοί, οι διοικητικοί υπάλληλοι, οι δικηγόροι κ.λ.π. Η λήψη και εφαρμογή μιας απόφασης, σε αρκετές περιπτώσεις δεν παραμένει εντός των στενών ορίων μιας επιχείρησης ή ενός οργανισμού, αλλά είναι δυνατόν να επηρεάζει και άλλους τομείς οικονομικούς και κοινωνικούς.

Η διαδικασία λήψης της απόφασης περιλαμβάνει εναλλακτικές αποφάσεις, οι οποίες τίθενται υπό εξέταση προκειμένου να επιλεγθεί η λύση που σύμφωνα με την πρόβλεψη αποτελεσμάτων, θα επιφέρει τα καλύτερα δυνατά αποτελέσματα με βάση τα δεδομένα του προβλήματος.

Σύμφωνα, με τον εμπνευστή της Ιεραρχικής Ανάλυσης Αποφάσεων (AHP), Saaty, σε δημοσίευσή του, το 1977 «... Η Ιεραρχική Ανάλυση Αποφάσεων αποτελεί την πιο ισχυρή μέθοδο ταξινόμησης που χρησιμοποιείται από το ανθρώπινο μυαλό για να διατάσσει παρατηρήσεις, οντότητες, εμπειρίες και πληροφορίες. Αν και ακόμη δεν έχει οριστεί από τη νευρολογία και την ψυχολογία, η ιεραρχική ταξινόμηση πιθανόν να αντιπροσωπεύει την πρωταρχική λειτουργία του συντονισμού και της οργάνωσης. Η χρήση της ιεραρχικής διάταξης πρέπει να είναι όσο παλιά είναι η ανθρώπινη σκέψη, συνειδητά και ασυνείδητα»

Προκειμένου να είναι δυνατή η μοντελοποίηση τέτοιου είδους προβλημάτων καθοριστικής σημασίας είναι οι σχετικές μετρήσεις. Οι μετρήσιμες ποσότητες λαμβάνονται σε περιπτώσεις που αντιστοιχούν σε οικονομικά μεγέθη ή δείκτες, ενώ σε περιπτώσεις που οι υπό μελέτη ποσότητες αφορούν την ποιότητα του περιβάλλοντος, την υγεία, ή ζητήματα που άπτονται κοινωνικών φαινομένων δεν δύναται η λήψη δεδομένων.

Αυτό που πραγματοποιείται με την Ιεραρχική Ανάλυση Αποφάσεων είναι η βαθμολογία μιας παραμέτρου σε σύγκριση με τις υπόλοιπες. Η διαδικασία στηρίζεται

σε δυαδικές συγκρίσεις, που σύμφωνα με την κλίμακα του Saaty παρέχουν και το αντίστοιχο αποτέλεσμα. Σημαντικό σημείο αποτελεί ο έλεγχος συνέπειας καθώς διαδραματίζει καθοριστικό ρόλο στην αποδοχή του αποτελέσματος, ενώ η δυσκολία αυξάνεται λόγω της υποκειμενικότητας στην κρίση του συνόλου.

Η Ιεραρχική Ανάλυση Αποφάσεων έχει χρησιμοποιηθεί σε πληθώρα περιπτώσεων όπως:

1. Στην IBM, προκειμένου να αναβαθμιστεί η ποιότητα του υπολογιστή AS/400. Η IBM για το συγκεκριμένο μοντέλο κέρδισε το διεθνές βραβείο Malcolm Baldrige National Quality Award.
2. Το 1986, χρησιμοποιήθηκε από το Ινστιτούτο Στρατηγικών και Αμυντικών Μελετών, κυβερνητικός οργανισμός, προκειμένου να προταθούν λύσεις για την κατάσταση όπως αυτή μαινόταν στη Ν. Αφρική, όπου επικρατούσε αστάθεια και συγκρούσεις.
3. Το 1987, η ίδια μέθοδος χρησιμοποιήθηκε προκειμένου να επιλεγεί η καταλληλότερη πλατφόρμα εξόρυξης πετρελαίου κοιτασμάτων από τον Ατλαντικό. Προκειμένου να ληφθεί η συγκεκριμένη απόφαση ο κύριος παράγοντας που λήφθηκε υπόψιν είναι το κόστος κατεδάφισης και όχι το κόστος υλοποίησης της πλατφόρμας.
4. Η μέθοδος χρησιμοποιήθηκε στη διαμάχη ανάμεσα σε Ηνωμένες Πολιτείες και Κίνα, το 1995 σχετικά με την εγγραφή πειρατικών δίσκων μουσικής, οπότε από με την μέθοδο αξιολογήθηκαν τα τυχόν πλεονεκτήματα, το κόστος και το ρίσκο. Από τα συμπεράσματα οι ΗΠΑ δεν προχώρησαν σε κυρώσεις προς την Κίνα, αντιθέτως την βράβευσαν για την επιχειρηματική της δραστηριότητα.
5. Το 1998, η μέθοδος χρησιμοποιήθηκε από την πολιτική αεροπορία της Βρετανίας, προκειμένου να επιλεγεί ο τρόπος ψυχαγωγίας των επιβατών κατά την διάρκεια της πτήσης.
6. Το 1999 η εταιρεία XEROX, με την Ιεραρχική Ανάλυση αποφάσεων κλήθηκε να επιλέξει σχετικά με την διάθεση 1.000.000.000 δολαρίων σε κάποιο ερευνητικό πρόγραμμα για το οποίο εκδήλωνε ενδιαφέρον εκείνη την εποχή.

7. Το 1999 η αυτοκινητοβιομηχανία Ford χρησιμοποίησε την εν λόγω μέθοδο με σκοπό την ιεράρχηση προτεραιοτήτων ώστε οι πελάτες της να ικανοποιούνται στο βέλτιστο βαθμό, ενώ στη συγκεκριμένη περίπτωση το λογισμικό που χρησιμοποιήθηκε βραβεύτηκε για την αποτελεσματικότητά του.

8. Το 2001, έγινε αξιολόγηση παραγόντων στην Τουρκία, προκειμένου να πραγματοποιηθεί η μετεγκατάσταση της τουρκικής πόλης Adapazarı, η οποία καταστράφηκε μετά από σεισμό.

Πέρα από τις παραπάνω αναφορές εφαρμογής της μεθόδου, η μέθοδος χρησιμοποιείται στον τομέα εργατικού δυναμικού των επιχειρήσεων στον κλάδο επιλογής, αξιολόγησης και προαγωγής του προσωπικού. Επίσης, στις αθλητικές ομάδες για την επιλογή παικτών και για την ανανέωση της συνεργασίας των παικτών με την ομάδα, επιπλέον, σε πλήθος κυβερνητικών και στρατιωτικών ζητημάτων. Από τα παραπάνω, γίνεται αντιληπτό ότι οι εφαρμογές της μεθόδου είναι πάρα πολλές και για τον λόγο αυτό και η έρευνα που πραγματοποιείται για την εν λόγω μέθοδο ιδιαίτερα ευρεία και μεγάλη.

Το πρώτο βήμα εφαρμογής της μεθόδου είναι ο διαχωρισμός του προβλήματος σε μικρότερα. Το δεύτερο βήμα περιλαμβάνει δυαδικές συγκρίσεις που αποσκοπούν στον καθορισμό της προτεραιότητας και την ιεραρχική τους τοποθέτηση. Οι τρεις αρχές στις οποίες στηρίζεται η μέθοδος είναι:

A) Η αποσύνθεση, διαδικασία στην οποία αναζητούνται βασικά στοιχεία του προβλήματος προκειμένου να συσταθεί μια ιεραρχία. Στη συνέχεια, η ανάλυση του προβλήματος λαμβάνει την μορφή δέντρου. Στο πρώτο επίπεδο, που βρίσκεται στο ανώτερο σημείο τοποθετείται ο τελικός στόχος, που αντιστοιχεί στην απόφαση. Στο δεύτερο επίπεδο τοποθετούνται τα κυριότερα κριτήρια λήψης της απόφασης, ενώ στο τελευταίο επίπεδο παρατίθενται οι εναλλακτικές αποφάσεις. Με αυτόν τον τρόπο, το πρόβλημα χωρίζεται σε μικρότερα τμήματα, ώστε να υπολογίζονται καλύτερα όλες οι παράμετροι και να γίνονται σαφείς.

B) Οι σχετικές συγκρίσεις, σε αυτό το δεύτερο μέρος της διαδικασίας πραγματοποιούνται συγκρίσεις ανά ζεύγη και για το κάθε κριτήριο αντιστοιχεί ένα

ποσοστό που καθορίζει και το πόσο σημαντικό είναι το εν λόγω κριτήριο. Οι συγκρίσεις πραγματοποιούνται ανά ζεύγη, για όλους τους πιθανούς συνδυασμούς ζευγών, κατά την διάρκεια της σύγκρισης όμως λαμβάνεται υπόψιν τα ζεύγη που βρίσκονται σε ανώτερα επίπεδα. Από την διαδικασία αυτή προκύπτει το ειδικό βάρος κάθε ζεύγους. Τέλος, εφαρμόζεται η αρχή της ιεραρχικής σύνθεσης όπου πολλαπλασιάζεται η σημαντικότητα κάθε στοιχείου, με το ζεύγος κατώτερου επιπέδου και το ζεύγος ανώτερου επιπέδου, οπότε και προκύπτει η τελική σημαντικότητα με σκοπό την λήψη της καλύτερης δυνατής απόφασης.

Γ) Η διάρθρωση των προτεραιοτήτων, που οδηγεί στην κατασκευή της ιεραρχίας. Ενώ σε κάθε προτίμηση αντιστοιχεί ένας συντελεστής βαρύτητας ώστε να κατασκευάζεται και ο ανάλογος πίνακας.

Πέρα από την δόμηση του προβλήματος εξίσου σημαντικός είναι ο υπολογισμός των κριτηρίων κάθε επιπέδου. Για το λόγο αυτό πρέπει να τεθούν προτεραιότητες ώστε να οριστούν οι συσχετισμοί των στοιχείων του κάθε επιπέδου. Για τα στοιχεία κάθε επιπέδου πραγματοποιείται σύγκριση, η οποία πραγματοποιείται εφόσον ληφθεί υπόψιν η σημαντικότητα κάθε στοιχείου ενώ η διαδικασία εκκινεί από την κορυφή και κατευθύνεται προς τα κάτω. Ο προσδιορισμός της σημαντικότητας προκύπτει από την σύγκριση ανά ζεύγη. Η διαδικασία στηρίζεται στην σύγκριση των κριτηρίων του δευτέρου επιπέδου της ιεραρχίας μεταξύ τους σε σχέση με το πόσο σημαντικά είναι. Σε ένα πρόβλημα με n πλήθος συνιστωσών, ο αριθμός συγκρίσεων που πραγματοποιούνται είναι $\frac{n(n-1)}{2}$.

Με την παραπάνω διαδικασία δημιουργείται ένα σύστημα το οποίο απεικονίζει έναν μικρόκοσμο. Μέσα από αυτό το σύστημα αξιολογούνται οι συνιστώσες και απαριθμούνται οι προτεραιότητες [5].

2.3.2 Ασαφής Λογική και Ιεραρχική Ανάλυση Αποφάσεων

Η ασαφής λογική δύναται να εξηγηθεί υπό δύο διαφορετικές οπτικές γωνίες. Υπό την στενή έννοια η ασαφής λογική αποτελεί ένα λογικό σύστημα, επέκταση της πολυσήμαντης λογικής. Υπό ευρύτερη έννοια, η οποία είναι και η πλέον χρησιμοποιούμενη, η ασαφής λογική είναι ταυτόσημη με την θεωρία των ασαφών

συνόλων. Στη θεωρία των ασαφών συνόλων, οι κλάσεις αντικειμένων δεν έχουν σαφή όρια και η συμμετοχή σε αυτές είναι θέμα βαθμού. Η επόμενη πρόταση συνοψίζει όλο το νόημα της ασαφούς θεωρίας: «Στη Ασαφή Λογική, η αλήθεια κάθε πρότασης είναι θέμα βαθμού». Τα χαρακτηριστικά της ασαφούς λογικής είναι τα παρακάτω:

1. Πρόκειται για θεωρία εύκολα κατανοητή.
2. Είναι ευέλικτη.
3. Είναι ανεκτική ακόμα και αν τα δεδομένα είναι ανακριβή.
4. Στην θεωρία της Ασαφούς Λογικής είναι δυνατή η μοντελοποίηση μη γραμμικών συναρτήσεων τυχαίας πολυπλοκότητας.
5. Η Ασαφής Λογική δομείται με βάση την εμπειρία των ειδικών.
6. Η Ασαφής Λογική ενσωματώνεται με «κανονικές» τεχνικές ελέγχου.
7. Η Ασαφής Λογική χρησιμοποιεί φυσική γλώσσα.

Στην Ασαφή Λογική μια μεταβλητή δεν περιγράφεται από αριθμούς αλλά από λέξεις, πρόκειται για την «γλωσσική μεταβλητή». Επίσης, στις περισσότερες εφαρμογές εφαρμόζεται ο κανόνας εάν – τότε, if – then, πρόκειται για την εφαρμογή του «ασαφή κανόνα». Στόχος της ασαφούς λογικής είναι η διερεύνηση των μεταβλητών εισόδου σε έναν χώρο εξόδου και πραγματοποιείται με μια λίστα προτάσεων που περιλαμβάνουν προτάσεις στις οποίες έχει εφαρμοστεί ο κανόνας if-then. Η αξιολόγηση των προτάσεων πραγματοποιείται παράλληλα ενώ η σειρά αξιολόγησης δεν διαδραματίζει ιδιαίτερο ρόλο.

Ασαφής Συμπερασμός

Ο ασαφής συμπερασμός στηρίζεται στην ερμηνεία των τιμών διανυσμάτων εισόδου και βάση ορισμένων κανόνων, ο ορισμός των τιμών διανυσμάτων εξόδου. Το σύνολο αυτών των τιμών συνιστά το ασαφές σύνολο, το οποίο προσδιορίζεται από την λέξη «ασαφές» καθώς δεν έχει σαφή όρια ενώ είναι δυνατόν να περιλαμβάνει και στοιχεία με μερικό βαθμό συμμετοχής. Γίνεται κατανοητό, ότι το ασαφές σύνολο συνιστά την

επέκταση της έννοιας του κλασικού συνόλου. Έστω Z ο χώρος εισόδου και τα στοιχεία αυτού z , τότε το ασαφές σύνολο C στο Z , ορίζεται από το σύνολο των διατεταγμένων στοιχείων:

$$C = \{z, m_C(z) | z \in Z\}$$

Ο βαθμός συμμετοχής καθορίζεται από την συνάρτηση συμμετοχής του z στο C , $m_C(z)$. Η συνάρτηση συμμετοχής παριστάνεται από μια καμπύλη που καθορίζει τον τρόπο με τον οποίο κάθε σημείο εισόδου αντιστοιχεί σε μια τιμή συμμετοχής ανάμεσα στο 0 και στο 1. Οι συναρτήσεις συμμετοχής για ασαφή σύνολα παριστάνονται με τριγωνικές, τραπεζοειδείς, Gaussian, συναρτήσεις καμπάνας κ.λ.π.

Η μέθοδος της Ιεραρχικής Ανάλυσης και οι παραλλαγές αυτής συνιστούν σημαντικό μέρος της μεθοδολογίας αποφάσεων που χρησιμοποιείται ευρέως. Οι λόγοι για τους οποίους η μέθοδος είναι τόσο διαδεδομένη οφείλεται πρώτον, στην δυνατότητα επεξεργασίας δεδομένων τα οποία εμφανίζουν υψηλό βαθμό αβεβαιότητας και ανακρίβειας, δεύτερον, η ευχρηστία της μεθόδου κατά την επίλυση προβλημάτων κατάταξης, τρίτον, η μαθηματική ακρίβεια και καθαρότητά της και τέταρτον, η διαφάνεια της ασαφούς λογικής και της θεωρίας ασαφών συνόλων.

Οι βασικές ιδιότητες των συναρτήσεων συμμετοχής είναι:

- A) Τα ασαφή σύνολα περιλαμβάνουν ασαφείς μεταβλητές (π.χ. κρύος καιρός)
- B) Τα ασαφή σύνολο αποδέχονται την πιθανότητα μερικής συμμετοχής σε αυτά (π.χ. Ο καιρός είναι αρκετά κρύος)
- Γ) Το κατά πόσο μια μεταβλητή αποτελεί μέρος του ασαφούς συνόλου, καθορίζεται από την τιμή της συνάρτησης συμμετοχής η οποία βρίσκεται ανάμεσα στο 0 και στο 1.
- Δ) Η συνάρτηση συμμετοχής ενός ασαφούς συνόλου απεικονίζει την τιμή εισόδου στην τιμή συμμετοχής. [6]

2.4 Μεθοδολογία Εξόρυξης Κανόνων

Η μεθοδολογία εξόρυξης κανόνων αφορά την ανακάλυψη των κανόνων συσχέτισης. Ένα από τα πιο σημαντικά σημεία στην εξόρυξη δεδομένων ενός συστήματος προκειμένου να είναι δυνατή η μελέτη του αποτελεί η ανακάλυψη κανόνων συσχέτισης. Στην πραγματικότητα, θεωρείται από πολλούς ως ο θεμέλιος λίθος της εξόρυξης δεδομένων, καθώς οι διαδικασίες, οι μεθοδολογίες και οι τεχνικές εξόρυξης προέρχονται κυρίως από μηχανική μάθηση, στατιστικές, βάσεις δεδομένων κ.λπ. Οι Κανόνες Συσχέτισης είναι ένα σύνολο κανόνων που χρησιμοποιούνται προκειμένου να είναι δυνατή η μελέτη και ο συσχετισμός των δεδομένων. Σκοπός της μεθοδολογίας είναι η ανακάλυψη σχέσεων που υφίσταται ανάμεσα στα δεδομένα. Αυτές οι σχέσεις προκύπτουν όταν οι τιμές κάποιων δεδομένων εμφανίζονται επανειλημμένα. Οι κανόνες συσχέτισης κάνουν εκτενή χρήση εννοιών, οι οποίες προέρχονται από τη θεωρία συνόλων. Για το λόγο αυτό κρίνεται χρήσιμο να προϋπάρχουν γνώσεις που αφορούν το υποσύνολο, την πράξη της τομής και της ένωσης καθώς και άλλους όρους που σχετίζονται με υποσύνολα, τομές και ενώσεις. Η δημιουργία κανόνων προϋποθέτει την διαδικασία της διακριτοποίησης, στην οποία περιλαμβάνονται διάφορες μέθοδοι όπως η μέθοδος Ανάλυσης Cluster, η μέθοδος k-Means ή η μέθοδος Self-Organizing Maps.

Οι κανόνες συσχέτισης είναι μια κοινή μέθοδος ανάλυσης που εφαρμόζεται κατά κόρον στα καλάθια αγορών, τόσο των φυσικών όσο και των ηλεκτρονικών καταστημάτων και αφορά την λήψη δεδομένων σχετικά με τον τρόπο που οι πελάτες του καταστήματος πραγματοποιούν τις αγορές. Κάθε αγορά αποτελείται από ένα σύνολο προϊόντων. Τα σύνολα πωλήσεων προϊόντων μπορούν να αναλυθούν προκειμένου να φανεί ποιες ομάδες προϊόντων πωλούνται συχνά μαζί. Από αυτές τις ομάδες υπολογίζονται κανόνες της μορφής «αν ένας πελάτης αγοράσει το προϊόν Α, έχει 70% πιθανότητα να αγοράσει ταυτόχρονα και το προϊόν Β, ενώ η ταυτόχρονη πώληση των προϊόντων Α και Β παρουσιάζεται στο 10% των συνολικών πωλήσεων». Αυτές οι πληροφορίες είναι χρήσιμες για τις εταιρείες. Από τα αποτελέσματα είναι πιθανόν να καθοριστεί η διάταξη και τοποθέτηση προϊόντων στο κατάστημα. Με την ύπαρξη δύο προϊόντων, που συχνά πωλούνται μαζί, σε παρακείμενες τοποθεσίες, επιτυγχάνεται αύξηση των πωλήσεων καθώς είναι πιθανόν ο πελάτης κατά την

διάρκεια των αγορών του να αγοράσει σχετικά προϊόντα εάν το ένα προϊόν βρίσκεται κοντά με το άλλο. Τα αποτελέσματα που προκύπτουν χρησιμοποιούνται επίσης στον προγραμματισμό προσφορών. Ο πωλητής, προσφέροντας χαμηλή τιμή για το προϊόν A, προσελκύει πελάτες και αυξάνει τις πωλήσεις του προϊόντος. Εάν τοποθετήσει το σχετικό προϊόν B σε παρακείμενα ράφια, θα αυξήσει τις πωλήσεις B, καθώς οι πελάτες είναι πιο πιθανό να αγοράσουν περισσότερα προϊόντα όταν δουν ότι οι τιμές είναι πιο προσιτές. Εφόσον κάποιος έχει αγοράσει ένα προϊόν A, με τη σειρά του, είναι πιο πιθανό να αγοράσει ένα άλλο προϊόν. Αυτές οι τεχνικές σταυροειδών πωλήσεων ισχύουν επίσης για την ταχεία πώληση προϊόντων με σύντομες ημερομηνίες λήξεως.

Έτσι, για την ανάλυση των δεδομένων έγινε αποθήκευση αυτών σε m στήλες δεδομένων. Οι τιμές των πεδίων είναι δυαδικές, 0 και 1. Έτσι, όταν συμβαίνουν ταυτόχρονα τα δεδομένα m_4 και m_{10} , τότε εισάγεται το 1 στην αντίστοιχη σειρά ενώ σε όλα τα υπόλοιπα δεδομένα εισάγεται 0. Από την ανάλυση δεδομένων προκύπτουν στοιχεία συσχετισμού για την κοινή εμφάνιση δεδομένων του συνόλου. Η δυσκολία του προβλήματος έγκειται στο πλήθος των συνδυασμών, καθώς σε 100 μεταβλητές αντιστοιχούν 250.000.000 κανόνες. Πλήρης έλεγχος αυτών των κανόνων σε μια βάση δεδομένων είναι εξαιρετικά χρονοβόρα.

Για την αντιμετώπιση αυτού του προβλήματος οι Agrawal, Imielinski & Swami (1993) πρότειναν τους Κανόνες Συσχέτισης. Οι κανόνες συσχέτισης έχουν την μορφή:

$$X \rightarrow Y, \text{ όπου } X \subseteq A \text{ και } Y \subseteq A \text{ ενώ } X \cap Y = \emptyset$$

Έτσι, αν το σύνολο A του υπό μελέτη συστήματος αποτελείται από 100 μεταβλητές $\{A_1, A_2, \dots, A_{100}\}$ το υποσύνολο X και το υποσύνολο Y δεν έχουν κοινά στοιχεία ενώ οι κανόνες συσχέτισης συνδέουν την εμφάνιση του A_{11} με το στοιχείο A_{102} , για παράδειγμα.

Το κατά πόσο ο κανόνας είναι ισχυρός, καθορίζεται από δύο ποσοτικά μεγέθη, τα οποία είναι η υποστήριξη (support) και η εμπιστοσύνη (confidence).

Η υποστήριξη αφορά τον αριθμό των συναλλαγών $X \rightarrow Y$ και εκφράζεται με την μαθηματική σχέση:

$$\mathit{supp}(X \rightarrow Y) = XUY$$

Η εμπιστοσύνη είναι η πιθανότητα εμφάνισης του X ταυτόχρονα με το Y , η μαθηματική σχέση που εκφράζει την εμπιστοσύνη είναι:

$$\mathit{conf}(X \rightarrow Y) = P(Y|X)$$

Για την δόμηση Κανόνων Συσχέτισης, ο χρήστης ορίζει ελάχιστες τιμές που αφορούν την υποστήριξη και την εμπιστοσύνη. Στη συνέχεια, ο αλγόριθμος τρέχει μέσω της βάσης δεδομένων, αναλύει τα δεδομένα και προσδιορίζει όλους τους κανόνες που έχουν υποστήριξη και εμπιστοσύνη ίση ή μεγαλύτερη από τις προεπιλεγμένες τιμές. Αυτοί οι κανόνες γενικά πιστεύεται ότι είναι έγκυροι. Επιπλέον, ο χρήστης ορίζει:

A) Το στοιχειοσύνολο, που περιέχει n στοιχεία, τα οποία εμφανίζονται μαζί., πχ. $I = \{A_5, A_{10}, A_{15}\}$

B) Την συχνότητα εμφάνισης ενός στοιχειοσυνόλου, που αφορά το πόσο συχνά εμφανίζονται τα εν λόγω στοιχεία.

Γ) Η υποστήριξη του στοιχειοσυνόλου που αφορά το ποσοστό συναλλαγών εντός του στοιχειοσυνόλου.

Δ) Συχνό στοιχειοσύνολο, που είναι εκείνο το στοιχειοσύνολο που απαντάται με συχνότητα μεγαλύτερη από μια ελάχιστη συχνότητα που έχει ορίσει ο χρήστης.

Δύο μέτρα ενός έγκυρου κανόνα είναι η υποστήριξη και η εμπιστοσύνη. Η υποστήριξη διασφαλίζει ότι ο κανόνας καλύπτει ένα ικανοποιητικό ποσοστό συναλλαγών. Οι κανόνες με χαμηλό βαθμό υποστήριξης μπορούν να θεωρηθούν ότι εκφράζουν ένα τυχαίο γεγονός. Η αυτοπεποίθηση είναι επίσης ένα μέτρο για το εάν η εμφάνιση της αριστερής πλευράς ανταμείβει την εμφάνιση του δεξιού μέρους του κανόνα. Οι άνθρωποι χρησιμοποιούν τις έννοιες της υποστήριξης και εμπιστοσύνης για να εξάγουν τους κανόνες. Ωστόσο, πρέπει να σημειωθεί ότι τα υψηλά επίπεδα

κοινωνικής υποστήριξης και εμπιστοσύνης δεν αποτελούν πάντα εγγύηση ότι ο κανόνας αναδεικνύει μια πραγματική σχέση.

2.4.1 Κανόνες Συσχέτισης

Ο μεγάλος όγκος δεδομένων δεν διευκολύνει την δόμηση κανόνων συσχέτισης. Η πιο απλή διαδικασία περιλαμβάνει την δημιουργία των περισσότερων δυνατών κανόνων, στη συνέχεια πρέπει να πραγματοποιηθεί ο υπολογισμός της υποστήριξης και της εμπιστοσύνης για κάθε έναν από τους κανόνες αυτούς και στο τελικό στάδιο να διατηρηθούν μόνο οι κανόνες που μέτρο μεγαλύτερο ή ίσο από μια ελάχιστη τιμή, που έχει οριστεί ως τιμή κατωφλίου. Εν τούτοις, η παραπάνω διαδικασία καθίσταται σχεδόν αδύνατη, καθώς για i πλήθος στοιχείων, το πλήθος των στοιχειοσυνόλων k δίνεται από την σχέση: $k = 2^i - 1$, έτσι για μόλις 20 στοιχεία ο αριθμός των δυνατών στοιχειοσυνόλων υπολογίζεται σύμφωνα με την παραπάνω σχέση σε 1.000.000. Σε ένα ρεαλιστικό πρόβλημα το i , είναι πιθανό να λαμβάνει τιμές της τάξεως των εκατοντάδων ή των χιλιάδων. Ως εκ τούτου, η εφαρμογή της παραπάνω διαδικασίας καθίσταται μη δυνατή.

Οι μέθοδοι που προτείνονται για την δημιουργία κανόνων συσχέτισης, περιλαμβάνει δύο στάδια:

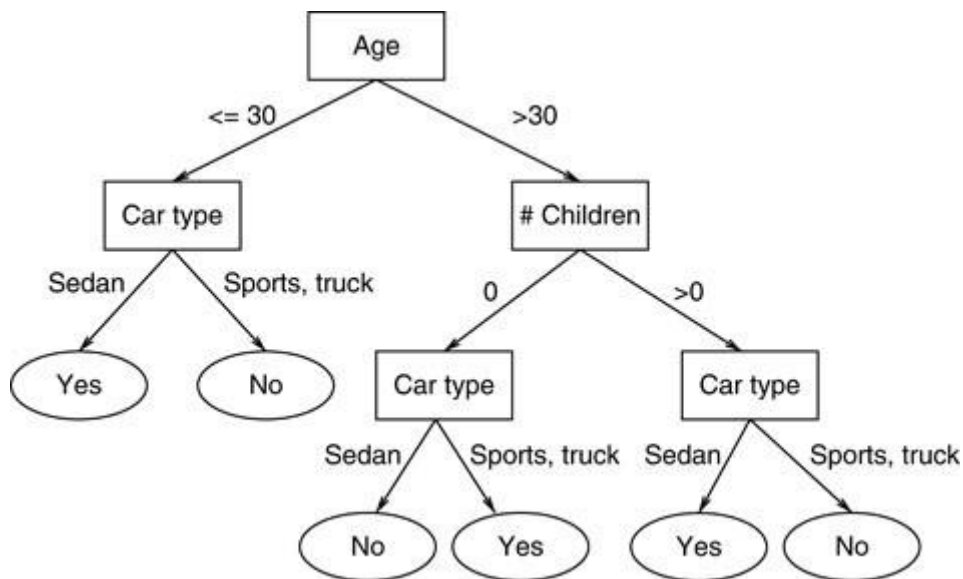
Στο πρώτο στάδιο, εντοπίζονται τα στοιχειοσύνολα με την μεγαλύτερη συχνότητα. Για τα στοιχειοσύνολα αυτά η υποστήριξη λαμβάνει τιμές μεγαλύτερες ή ίσες από την τιμή κατωφλίου. Το δεύτερο στάδιο περιλαμβάνει την δημιουργία κανόνων συσχέτισης που σχετίζονται κατά αποκλειστικότητα με τα στοιχειοσύνολα που εμφανίζονται με την μεγαλύτερη συχνότητα. Οι κανόνες που δημιουργούνται με αυτόν τον τρόπο πρέπει να ικανοποιούν την σχέση εμπιστοσύνης.

2.4.2 Δέντρα Αποφάσεων

Στην εξόρυξη δεδομένων, ένα δέντρο αποφάσεων είναι ένα μοντέλο πρόβλεψης που μπορεί να χρησιμοποιηθεί για την αναπαράσταση ταξινομητών και μοντέλων για παλινδρόμηση. Στην επιχειρηματική έρευνα, τα δέντρα αποφάσεων αναφέρονται σε ένα μοντέλο ιεραρχίας αποφάσεων, καθώς και στις συνέπειές τους.

Το άτομο που παίρνει την απόφαση ακολουθεί τα δέντρα αποφάσεων για να βρει την καλύτερη στρατηγική για να πετύχει τον στόχο του. Όταν ένα δέντρο αποφάσεων χρησιμοποιείται για την κατάταξη διεργασιών, είναι πιο κατάλληλο να γίνεται αναφορά σε αυτό ως δέντρο ταξινόμησης. Όταν χρησιμοποιείται για διαδικασίες παλινδρόμησης, προτιμάται να ονομάζεται δέντρο παλινδρόμησης.

Το δέντρο ταξινόμησης χρησιμοποιείται για την ταξινόμηση ενός αντικειμένου ή ενός συμβάντος. Ένα παράδειγμα συμβάντος είναι ένας ασφαλισμένος του ιδιωτικού ή δημόσιου τομέα ενώ ο εν λόγω ασφαλισμένος εντάσσεται και σε ένα σύνολο προκαθορισμένων κατηγοριών (π.χ. επικίνδυνη κατηγορία/μη επικίνδυνη κατηγορία). Οι παράμετροι αυτοί στηρίζονται στη χαρακτηριστική αξία κάθε παραμέτρου, όπως η ηλικία ή το φύλο. Τα δέντρα χρησιμοποιούνται συχνά σε τομείς όπως η οικονομία, το μάρκετινγκ, η μηχανική και η ιατρική. Το δέντρο ταξινόμησης είναι χρήσιμο για την εξερεύνηση σχέσεων μεταξύ δεδομένων. Εν τούτοις, τα δέντρα αποφάσεων δεν αποσκοπούν στην αντικατάσταση των ήδη εφαρμοζόμενων παραδοσιακών μεθόδων στατιστικής ενώ υπάρχουν πολλές άλλες τεχνικές που μπορούν να χρησιμοποιηθούν για ταξινόμηση ή για πρόβλεψη της συμπερίληψης γεγονότων σε προκαθορισμένες κατηγορίες συνόλου, όπως τεχνητά νευρωνικά δίκτυα ή μηχανές διανυσμάτων υποστήριξης.



Εικόνα 2: Τυπικό δέντρο αποφάσεων για την ταξινόμηση δεδομένων (Dobra, 2009)

Παραπάνω, στην Εικόνα 2 απεικονίζεται ένας τυπικός ταξινομητής δέντρου αποφάσεων. Ένα παράδειγμα δέντρου αποφάσεων χρησιμοποιείται για να βοηθήσει τους ανθρώπους να υποβάλουν αίτηση για τραπεζικό δάνειο. Ως μέρος αυτής της διαδικασίας, ο αιτών θα συμπληρώσει ως μέρος αυτής της διαδικασίας, ο αιτών συμπληρώνει μια φόρμα αίτησης. που περιλαμβάνει τον αριθμό των εξαρτώμενων (αναλογικά), την αναλογία «Αναλογία δανείου προς αξία» (LTV), την οικογενειακή κατάσταση (MARST - οικογενειακή κατάσταση), την αναλογία πληρωμής προς εισόδημα (PAYINC - αναλογία πληρωμής προς εισόδημα), το επιτόκιο, έτη σε τρέχουσα διεύθυνση (YRSADD - έτη στην τρέχουσα διεύθυνση) και έτη στην τρέχουσα εργασία (YRSJOB - έτη στην τρέχουσα εργασία).

Με βάση τα παραπάνω στοιχεία, ο ανάδοχος θα αποφασίσει εάν η αίτηση δανείου πρέπει να εγκριθεί ή όχι. Το δέντρο αποφάσεων βασίζεται στα πεδία που εμφανίζονται στις φόρμες αίτησης δανείου. Το παραπάνω παράδειγμα χρησιμοποιεί δέντρο αποφάσεων για να αναπαραστήσει ένα μοντέλο ταξινόμησης. Επιπλέον, συχνά χρησιμοποιείται και η διαδικασία εξαγωγής γνώσης, δηλαδή η απόκτηση γνώσης από έναν άνθρωπο ειδικό (ή ανθρώπινη τεχνογνωσία) για χρήση από ένα ευφυές σύστημα. Η εξαγωγή γνώσης είναι συνήθως μια δύσκολη διαδικασία, επειδή δεν είναι εύκολο να βρεθεί ένας διαθέσιμος εμπειρογνώμονας, ο οποίος είναι σε θέση, έχει το χρόνο και είναι πρόθυμος να παρέχει τις πληροφορίες που απαιτούνται για τη δημιουργία ενός αξιόπιστου έμπειρου συστήματος. Στην πραγματικότητα, η δυσκολία συνδέεται άμεσα με την διαδικασία και αυτός είναι ο λόγος που οι εταιρείες απορρίπτουν τα ευφυή συστήματα.

Η χρήση ενός δέντρου αποφάσεων είναι μια πολύ δημοφιλής τεχνική στην εξόρυξη δεδομένων που χρησιμοποιείται σε πολλά διαφορετικά πεδία. Σύμφωνα με πολλούς ερευνητές, τα δέντρα αποφάσεων είναι δημοφιλή λόγω της απλότητάς τους και της διαφάνειάς τους. Τα δέντρα απόφασης είναι αυτονόητα. ένας ειδικός εξόρυξης δεδομένων δεν χρειάζεται να είναι συγκεκριμένος τύπος για να ακολουθήσει μια διαδικασία δέντρου αποφάσεων. Τα περισσότερα δέντρα αντιπροσωπεύονται συνήθως με μια γραφική αναπαράσταση ενός επιπέδου, η οποία καθιστά ευκολότερη σε σχέση με διαφορετικές τεχνικές. Αν το δέντρο κατάταξης

γίνεται πολύπλοκο χρειάζονται διαφορετικές γραφικές διαδικασίες για να απλοποιηθεί η ερμηνεία των αποτελεσμάτων τους.

2.5 Συστήματα μάθησης

2.5.1 Συστήματα Πρόσβασης Πληροφοριών

Εδώ και δεκαετίες υπάρχουν πολλοί και διαφορετικοί τύποι πληροφοριακών συστημάτων, έτσι, έχουν αναπτυχθεί συστήματα διεκπεραίωσης, συστήματα συναλλαγών, συστήματα υποστήριξης, συστήματα αλγοριθμικής λήψης αποφάσεων, συστήματα διαχείρισης γνώσης, συστήματα μάθησης, συστήματα αυτοματισμού και συστήματα βάσεων δεδομένων. Τα εν λόγω πληροφοριακά συστήματα στηρίζουν τον σχεδιασμό τους στην τεχνολογική ανάπτυξη και σχεδιάζονται με σκοπό την εκτέλεση εργασιών αδύνατες για τα όρια του ανθρώπινου εγκεφάλου. Στις εργασίες αυτές συγκαταλέγονται η διαχείριση μεγάλου όγκου πληροφοριών, η εκτέλεση πολύπλοκων υπολογισμών και ο ταυτόχρονος και πολλαπλός έλεγχος πλήθους πληροφοριών.

Τα πληροφοριακά συστήματα έχουν ευρεία εφαρμογή σε πλήθος εταιρειών και οργανισμών ενώ έχουν δημιουργηθεί θέσεις εργασίας σχετικά με τη διαχείριση των Πληροφοριακών Συστημάτων και της Πληροφορικής που φέρουν ειδική βαρύτητα στις εν λόγω εταιρείες και τους οργανισμούς, κατά αντιστοιχία με την θέση του Διευθύνοντα Συμβούλου, του Οικονομικού Διευθυντή ή του Τεχνικού Διευθυντή. Επίσης, εξέχουσα σημασία παρουσιάζει και το ζήτημα της ασφάλειας πληροφοριών για το λόγο αυτό εταιρείες που διαχειρίζονται πλήθος προσωπικών δεδομένων και δεδομένων ευαίσθητου χαρακτήρα έχουν θέσεις Διευθυντών Ασφάλειας Πληροφοριών.

Κάθε πληροφοριακό σύστημα αποτελείται από:

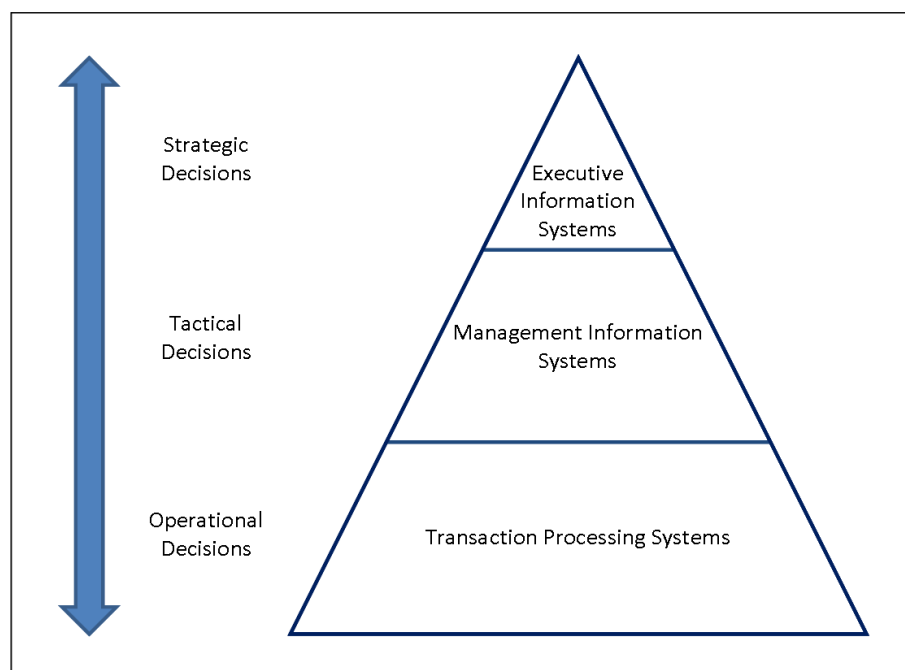
- i) το Υλικό Μέρος, το οποίο περιλαμβάνει την κεντρική μονάδα επεξεργασίας, τις συσκευές εισόδου – εξόδου, αποθήκευσης και επικοινωνίας.
- ii) το Λογισμικό Μέρος, το οποίο περιλαμβάνει το σύνολο των προγραμμάτων τα οποία απαιτούνται για την εύρυθμη λειτουργία του πληροφοριακού συστήματος
- iii) τα Δεδομένα, τα οποία αποθηκεύονται στις συσκευές αποθήκευσης

iv) οι διαδικασίες, δηλαδή ο προγραμματισμός που ακολουθείται προκειμένου να τελεστή η οποιαδήποτε ενέργεια

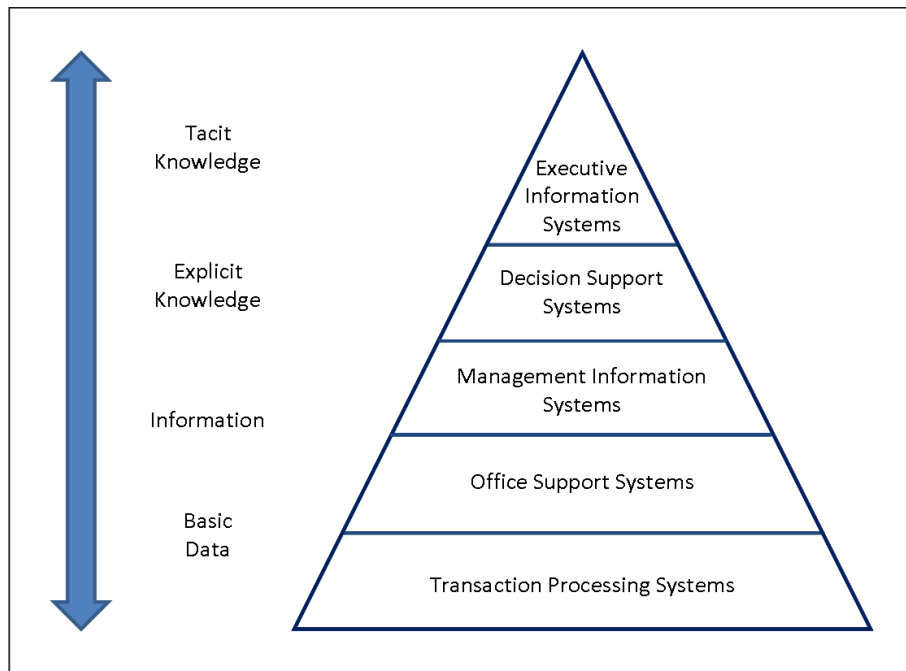
v) οι χειριστές του συστήματος. Οι χειριστές, το ανθρώπινο δυναμικό που χειρίζεται το εκάστοτε πληροφοριακό σύστημα καθορίζει σε μεγάλο βαθμό την επιτυχία ή αποτυχία του. Το τμήμα του ανθρώπινου δυναμικού πέραν από όσους εργάζονται για την χειρισμό του πληροφοριακού συστήματος περιλαμβάνει και τους χειριστές που σχετίζονται με την είσοδο και αποθήκευση των δεδομένων σε αυτό, καθώς αυτά τα δεδομένα σε έναν δεύτερο χρόνο θα υποστούν την απαραίτητη επεξεργασία.

vi) η ανατροφοδότηση, ουσιώδεις διαδικασία στον σύγχρονο πληροφοριακό κόσμο, που σχετίζεται με το κατά πόσο η αρχική πληροφορία λειτουργεί κατόπιν επεξεργασίας ως γεννήτορας νέων πληροφοριών που θα ανατροφοδοτήσουν το σύστημα και θα διαδραματίσουν ενεργό ρόλο στην υποστήριξη των τελικών αποφάσεων.

Κατά την δεκαετία του 1980, με βάση τα πληροφορικά συστήματα σχεδιάστηκε μια πυραμίδα οργάνωσης της εταιρείας ή του οργανισμού.



Εικόνα 3: Πληροφορικά συστήματα – μοντέλο πυραμίδας τριών επιπέδων (Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017)

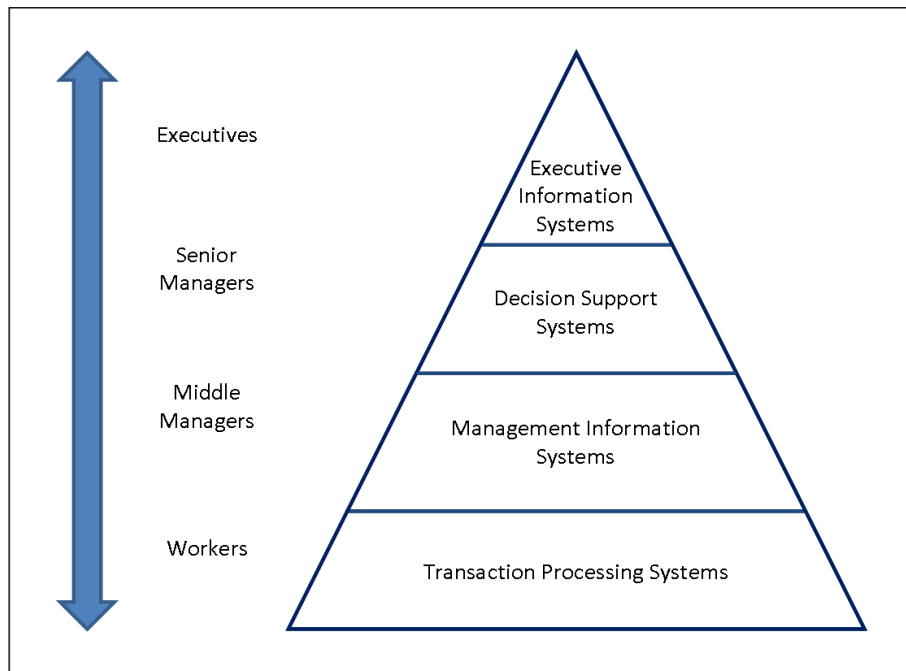


Εικόνα 4: Πληροφοριακά συστήματα – μοντέλο πυραμίδας πέντε επιπέδων (Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017)

Είναι ορατό ότι στη βάση της πυραμίδας τοποθετούνται τα συστήματα επεξεργασίας, ακολουθούν τα συστήματα διαχείρισης πληροφοριών, στη συνέχεια τοποθετούνται τα συστήματα λήψης αποφάσεων και στην κορυφή τα συστήματα υποστήριξης της διοίκησης. Το παραπάνω αρχικό μοντέλο της πυραμίδας, στο πέρασμα των ετών έχει υποστεί αρκετές τροποποιήσεις κατά περίπτωση. Εν τούτοις, η δημιουργία ολοένα και πιο προηγμένων τεχνολογικών πληροφοριακών συστημάτων δυσχεραίνει την ενσωμάτωσή τους σε αυτό.

Στα πιο προηγμένα πληροφοριακά συστήματα περιλαμβάνονται οι Αποθήκες Δεδομένων (Data Warehouses), Συστήματα Διαχείρισης Επιχειρηματικών Πόρων (ERP), τα Έμπειρα Συστήματα (Expert Systems), οι Μηχανές Αναζήτησης (Search Engines) και τα Γεωγραφικά Πληροφοριακά Συστήματα (GIS) [7].

Ανάλογα με το υιοθετούμενο μοντέλο πυραμίδας, διαφορετικοί εμπλεκόμενοι δραστηριοποιούνται σε κάθε διαφορετικό επίπεδο:



Εικόνα 5: Πληροφοριακά συστήματα – μοντέλο πυραμίδας πέντε επιπέδων (Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017)

2.5.1.2. Μηχανές Αναζήτησης

Ο παγκόσμιος ιστός αποτελεί μια τεράστια πηγή πληροφοριών τα τελευταία χρόνια. Αποτελείται από διασυνδεδεμένες ιστοσελίδες, οι οποίες προστίθενται συνεχώς σε αυτόν, με τρόπο που ως ώρας δεν είναι αυστηρά δομημένος. Ο όγκος των πληροφοριών που περιέχονται σε αυτό αυξάνεται γρήγορα και συνεχώς. Το Διαδίκτυο είναι ένας τεράστιος και περίπλοκος ιστός στον οποίο μπορούν να συμμετέχουν όλοι. Η δημιουργία προσωπικής ιστοσελίδας, ακόμη και χωρίς τη χρήση συγκεκριμένου προτύπου είναι εύκολη. Άνθρωποι λοιπόν με διαφορετικό μορφωτικό επίπεδο, κουλτούρα, γλώσσα και συχνά διαφορετικό σκοπό, καλό ή δόλιο, προσθέτουν συνεχώς νέο υλικό, εμπλουτίζοντας και αλλάζοντας την εικόνα του. Ο αριθμός νέων χρηστών του διαδικτύου που αναζητούν πληροφορίες σχετικά με τα θέματα που τους ενδιαφέρουν πληθαίνουν.

Για να υπολογίσουμε το μέγεθος του διαδικτύου, αρκεί να λάβουμε υπόψη ότι, ήδη, μέχρι το 1997, υπήρχαν περίπου 150.000.000 ιστοσελίδες. Το 2000, ο αριθμός των ιστοσελίδων έφτασε τα 300.000.000 περίπου, ενώ καθημερινά ο αριθμός τους αυξανόταν σε περίπου 1.000.000 ιστοσελίδες. Το 2002, ο αριθμός των ατόμων

με πρόσβαση στο διαδίκτυο έφτασε τα 8,9 δισεκατομμύρια. Το 2008, η Google ανέφερε ότι υπήρχαν 1.000.000.000.000.000 ιστοσελίδες, καταγεγραμμένες στην απογραφή εκείνου του έτους. Σήμερα ο αριθμός των ιστοσελίδων αυξάνεται με ρυθμό 4,5 εκατομμύρια ιστοσελίδες ανά μήνα. Ο τεράστιος όγκος πληροφοριών που βρίσκονται σήμερα στον Ιστό είναι κατανοητός εφόσον ληφθούν υπόψιν οι παραπάνω αριθμοί. Ο Ιστός είναι ένα μέρος όπου οι άνθρωποι μπορούν να βρουν πληροφορίες που σχετίζονται με αυτό που αναζητούν. Η διαρκής επέκταση του διαδικτύου δείχνει την εικόνα μιας άναρχης δομής. Για το λόγο αυτό η προσπάθεια αναζήτησης ορισμένων πληροφοριών, χαρακτηριζόταν από δυσκολία του χρήστη να εντοπίσει ιστοσελίδες που ταιριάζουν με το προς μελέτη αντικείμενο. Το τελευταίο ζήτημα είναι επίσης ένα σημαντικό πρόβλημα σε ερωτήματα ευρείας θεματολογίας, επίσης γνωστό ως το «πρόβλημα της αφθονίας». Το πρόβλημα της αφθονίας περιγράφει τον πολύ μεγάλο αριθμό ιστοσελίδων, με θέμα σχετικό με την αναζήτηση του χρήστη που οδηγεί σε αδυναμία επεξεργασίας και επιλογής από την πλευρά του χρήστη της κατάλληλης ιστοσελίδας. Το επόμενο ζήτημα που εντοπίζεται αφορά θέματα αναζήτησης μικρού εύρους, ζητήματα δηλαδή που είναι ιδιαίτερα στοχευμένα. Αυτή η περίπτωση περιγράφεται ως «Το πρόβλημα της έλλειψης», στην προκειμένη περίπτωση ο αριθμός των ιστοσελίδων που περιέχουν την επιθυμητή πληροφορία είναι μικρός και δύσκολα αναγνωρίσιμος.

Αυτές οι δυσκολίες οδήγησαν στην ανάγκη δημιουργίας εργαλείων που θα μπορούσαν να κάνουν αναζήτηση στο διαδίκτυο και να παρέχουν στον χρήστη τις πληροφορίες που αναζητούσε. Με τα συγκεκριμένα εργαλεία απλοποιείται και διευκολύνεται η αναζήτηση των επιθυμητών προϊόντων, θεμάτων, υπηρεσιών στον παγκόσμιο ιστό. Για να κατανοήσουμε πώς λειτουργούν οι μηχανές αναζήτησης, πρέπει πρώτα να προσεγγίσουμε δύο ζητήματα. Το πρώτο θέμα αφορά την κατανόηση της δομής του διαδικτύου και το δεύτερο τον τρόπο αναζήτησης από την πλευρά των χρηστών.

Το διαδίκτυο επεκτείνεται συνεχώς, με νέες ιστοσελίδες να προστίθενται σε τακτική βάση. Ωστόσο, οι υπάρχουσες ιστοσελίδες ανανεώνουν συχνά την δομή τους. Οι πληροφορίες που παρουσιάζονται σε αυτές τις ιστοσελίδες είναι αρκετά

συγκεχυμένες, καθώς για παράδειγμα, άλλες σελίδες είναι σχολαστικά βυθισμένες σε ένα θέμα, ενώ άλλες για το ίδιο θέμα, απλώς κάνουν μια υπόδειξη. Ωστόσο, σε πολλές περιπτώσεις, οι πληροφορίες είναι καθοδηγούμενες. Οι παρεχόμενες πληροφορίες ενδέχεται να χρησιμοποιηθούν για συγκεκριμένους διαφημιστικούς σκοπούς και ως εκ τούτου ενδέχεται να μην είναι απολύτως έγκυρες.

Οι μηχανές αναζήτησης λειτουργούν με αλγορίθμους αναζήτησης, προκειμένου να έχουν πρόσβαση σε πληροφορίες που έχουν αποθηκευτεί σε μία δομή δεδομένων. Οι αλγόριθμοι αναζήτησης είναι συγκεκριμένοι για κάθε μηχανή αναζήτησης και εξαρτώνται από το μέγεθος και την δομή των δεδομένων τους ώστε να δύναται να καθορίσουν την κατάταξη των ιστοσελίδων στα αποτελέσματα αναζήτησης. Οι αλγόριθμοι που χρησιμοποιούνται στις μηχανές αναζήτησης είναι:

A) Γραμμικοί Αλγόριθμοι: Οι γραμμικοί αλγόριθμοι είναι οι πιο βασικοί από όλους τους αλγορίθμους καθώς η εφαρμογή τους απαιτεί χρήση ελάχιστων κωδικών. Η χρήση τους ενδείκνυται σε περιπτώσεις που οι λίστες δεν είναι ταξινομημένες και τα δεδομένα έχουν αποθηκευτεί με τυχαίο τρόπο, καθώς εξετάζει τα στοιχεία ως λίστα είναι γενικά αναποτελεσματικός και η χρήση του δεν ενδείκνυται για τις μηχανές αναζήτησης.

B) Δυαδικός αλγόριθμος αναζήτησης: Σε αντίθεση με τους γραμμικούς αλγορίθμους, ο δυαδικός αλγόριθμος εκμεταλλεύεται τη σειρά μιας λίστας. Η χρήση δυαδικών αλγορίθμων ενδείκνυται στις περιπτώσεις που οι λίστες εμφανίζονται κατά σειρά αυξανόμενου μεγέθους. Έτσι, αν η αναζήτηση περιορίζεται σε όρους που βρίσκονται από το μέσον και κάτω, το επάνω μισό της λίστας εξαλείφεται. Αντίστοιχα, αν η αναζήτηση αναφέρεται στα μέγιστα της λίστας εξαλείφεται το κάτω μισό της λίστας. Η παραπάνω διαδικασία επαναλαμβάνεται συνεχώς εφόσον η δυαδική αναζήτηση χρησιμοποιεί ένα βρόχο ή πραγματοποιεί μια αναδρομή ώστε μετά από κάθε σύγκριση, ο χώρος να διαιρείται στο μισό. Οι δυαδικοί αλγόριθμοι είναι ιδιαίτερα αποτελεσματικοί σε περιπτώσεις που οι βάσεις δεδομένων είναι ιδιαίτερα εκτενείς, οπότε και είναι εντελώς αναποτελεσματικοί οι γραμμικοί αλγόριθμοι.

Οι αλγόριθμοι αναζητήσεων επηρεάζουν την κατάταξη της ιστοσελίδας στις μηχανές αναζήτησης. Κάθε μηχανή αναζήτησης χρησιμοποιεί ένα σύνολο κανόνων από τους οποίους «αντιλαμβάνεται» αν μια ιστοσελίδα είναι πραγματική ή spam. Από τα αποτελέσματα ελέγχου καθορίζεται η κατάταξη της ιστοσελίδας κατά την εμφάνιση των σχετικών με την αναζήτηση αποτελεσμάτων [19].

2.5.1.3 Ειδησεογραφία μέσω κοινωνικών μέσων

Ως μέσα κοινωνικής δικτύωσης λογίζονται όλα τα τεχνικά μέσα, τα οποία χρησιμεύουν στην συλλογή πληροφοριών και την επικοινωνία. Το κοινό στο οποίο απευθύνονται τα μέσα κοινωνικής δικτύωσης είναι ιδιαίτερα διευρυμένο λόγω της ευκολίας της χρήσης τους [14]. Μέχρι πριν κάποιες δεκαετίες, φορείς πληροφοριών ήταν ο έντυπος τύπος, το ραδιόφωνο και στη συνέχεια η τηλεόραση. Οι εν λόγω μορφές πληροφόρησης συνιστούν τους παραδοσιακούς τρόπους ενημέρωσης και αποτελούν πομπούς πληροφόρησης, οι οποίοι δημιουργούν δρόμους επικοινωνίας μονής κατεύθυνσης, υπό την έννοια ότι δεν υπήρχε αλληλεπίδραση μεταξύ των χρηστών, ούτε και η δυνατότητα ανταλλαγής απόψεων. Η έλευση του Web 2.0 άλλαξε το σκηνικό του διαδικτύου εφόσον προσέφερε την δυνατότητα αλληλεπίδρασης μεταξύ των χρηστών στα μέσα επικοινωνίας. Με τον τρόπο αυτό τα μέσα ενημέρωσης απέκτησαν κοινωνική χροιά, γεγονός που τα διαφοροποίησε από τα καθιερωμένα πρότυπα ενημέρωσης [15].

Σύμφωνα με τους Kaplan & Haenlein (2010), το πρώτο μέσον κοινωνικής δικτύωσης με την μορφή που τα γνωρίζουμε σήμερα ήταν το Open Diary, δύο δεκαετίες πίσω. Το Open Diary, αποτέλεσε μια πρώιμη ιστοσελίδα κοινωνικής δικτύωσης. Ιδρυτές της σελίδας ήταν οι Bruce και Susan Abelson και ο σκοπός της σελίδας ήταν η διαδικτυακή συγκέντρωση των συγγραφέων ημερολογίων. Την ίδια περίοδο χρησιμοποιήθηκε και ο όρος weblog, ο οποίος στη συνέχεια απλοποιήθηκε στον όρο blog. Το 2003, εμφανίστηκε το my space και έναν χρόνο αργότερα το Facebook, η δημοτικότητα του οποίου συνέβαλλε καθοριστικά στη θέση των μέσων κοινωνικής δικτύωσης. [16]

Το πλήθος των χρηστών του Facebook ενέπνευσε την ίδρυση παρόμοιων διαδικτυακών τόπων παρόμοιου περιεχομένου. Τα μέσα κοινωνικής δικτύωσης αυξάνονταν συνεχώς προσελκύοντας ολοένα και μεγαλύτερο πλήθος χρηστών, με αποτέλεσμα να συγκεντρώσουν το ενδιαφέρον του Marketing αλλά και την ακαδημαϊκής κοινότητας. Σύμφωνα με τον Evans (2008), τα μέσα κοινωνικής δικτύωσης αποτελούν την επιτομή του εκδημοκρατισμού της πληροφορίας, ενώ οι χρήστες τους αποκτούν έναν ενεργητικό ρόλο στην διαμόρφωση της πληροφορίας. Ταυτόχρονα, αποτελούν μέσα επικοινωνίας και ανταλλαγής απόψεων, παρέχουν κοινωνική και ψυχολογική υποστήριξη σε άτομα, ομάδες ή φορείς που βρίσκονται σε ανάγκη, ενώ τέλος αποτελούν για μεγάλο μέρος του πληθυσμού την σημαντικότερη πηγή πληροφόρησης.[17], [18]

Τα κοινωνικά δίκτυα και η προβολή ειδήσεων σχετίζονται με την χρήση αλγορίθμων. Από την έκθεση «New powers, new responsibilities. A global survey of journalism and artificial intelligence» που πραγματοποιήθηκε από το Polis του LSE και το Google News Initiative, οι δημοσιογραφικοί οργανισμοί που εκμεταλλεύονται τη χρήση τεχνητής νοημοσύνης εκμεταλλεύονται τρεις άξονες:

A) Συλλογή ειδήσεων: Συλλογή πληροφοριών, δημιουργία ιδεών για ιστορία, παρακολούθηση τάσεων, έρευνα, παρακολούθηση συμβάντων ή ζητημάτων, εξαγωγή πληροφοριών ή περιεχομένου.

B) Παραγωγή Ειδήσεων: Η παραγωγή ειδήσεων είναι η διαδικασία δημιουργίας και επεξεργασίας περιεχομένου για διαφορετικές μορφές και πλατφόρμες, δημιουργία κειμένου, εικόνας και βίντεο και επαναπροσδιορισμού περιεχομένου για διαφορετικούς τύπους κοινού.

Γ) Διανομή ειδήσεων: εξατομίκευση, μάρκετινγκ, εύρεση κοινού, κατανόηση της συμπεριφοράς των χρηστών, δημιουργία εσόδων / συνδρομές.

Σήμερα, οι μηχανές τεχνητής νοημοσύνης εκτελούν εργασίες σε διάφορους τομείς, για παράδειγμα οι δημοσιογράφοι του Bloomberg News χρησιμοποιούν αλγορίθμους τεχνητής νοημοσύνης που τους ειδοποιούν για τάσεις ή μοτίβα σε μεγάλες βάσεις δεδομένων.

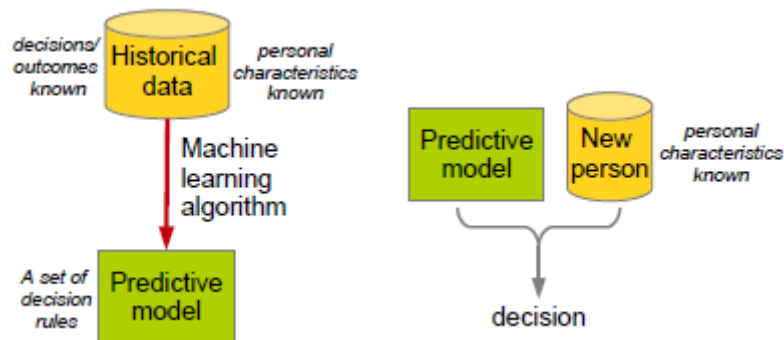
Το AFP (Agence France-Presse) χρησιμοποιεί μια εφαρμογή για να ανιχνεύει «πειραγμένες» φωτογραφίες, καθώς τα ρομπότ τεχνητής νοημοσύνης μπορούν να αναλύσουν πολύπλοκα δεδομένα σε σύντομο χρονικό διάστημα. Εδώ οι δημοσιογράφοι κάνουν σημαντική δουλειά, ελέγχοντας και αναλύοντας γεγονότα και επίσης συλλέγοντας πληροφορίες. Σε αυτό το πλαίσιο Ο δημοσιογράφος βρίσκεται στο επίκεντρο της διαδικασίας του ρεπορτάζ και είναι αναντικατάστατος. Τόσο το AFP αλλά και το Reuters χρησιμοποιούν συστήματα εντοπισμού ψεύτικων βίντεο και ψευδών ειδήσεων, τα deepfakes και Fake news. Η τελική αξιολόγηση, εν τούτοις προκύπτει από τους δημοσιογράφους καθώς τα στοιχεία πριν προβληθούν ή απορριφθούν πρέπει να ελέγχονται από πολλαπλές πηγές.

2.6 Δίκαιο κατά την αλγοριθμική λήψη αποφάσεων

Τι είναι όμως το δίκαιο κατά την αλγοριθμική λήψη αποφάσεων; στα πλαίσια της αλγοριθμικής λήψης αποφάσεων, της κατασκευής προβλεπτικών μοντέλων, της εξόρυξης δεδομένων και της ταξινόμησης των δεδομένων θα πρέπει να διασφαλίζονται Οι παρακάτω συνθήκες προκειμένου να είναι δίκαιη η διαδικασία μηχανικής (αλγοριθμικής) λήψης απόφασης (Fukuchi, et al., 2015):

- Τα άτομα σε έναν πληθυσμό τα οποία έχουν τα ίδια μη προστατευόμενα χαρακτηριστικά θα πρέπει να γίνονται αποδέκτες παρόμοιων αποφάσεων
- Τυχόν διαφορές στις αποφάσεις ανάμεσα σε διαφορετικούς ανθρώπους για ανάμεσα σε διαφορετικές ομάδες ανθρώπων θα πρέπει να εξηγούνται από παράγοντες άλλους από τα μη προστατευόμενα χαρακτηριστικά

Οι παραπάνω συνθήκες είναι γνωστές και ως συνθήκες Lipschitz ή συνθήκες στατιστικής ισοτιμίας, όπως προτάθηκαν για πρώτη φορά το 2012 (Dwork, et al., 2012).



Εικόνα 6: Τυπική δομή αλγοριθμικής λήψης αποφάσεων με γνωστά τα δημογραφικά και άλλα προστατενόμενα χαρακτηριστικά (Žliobaitė, 2017a)

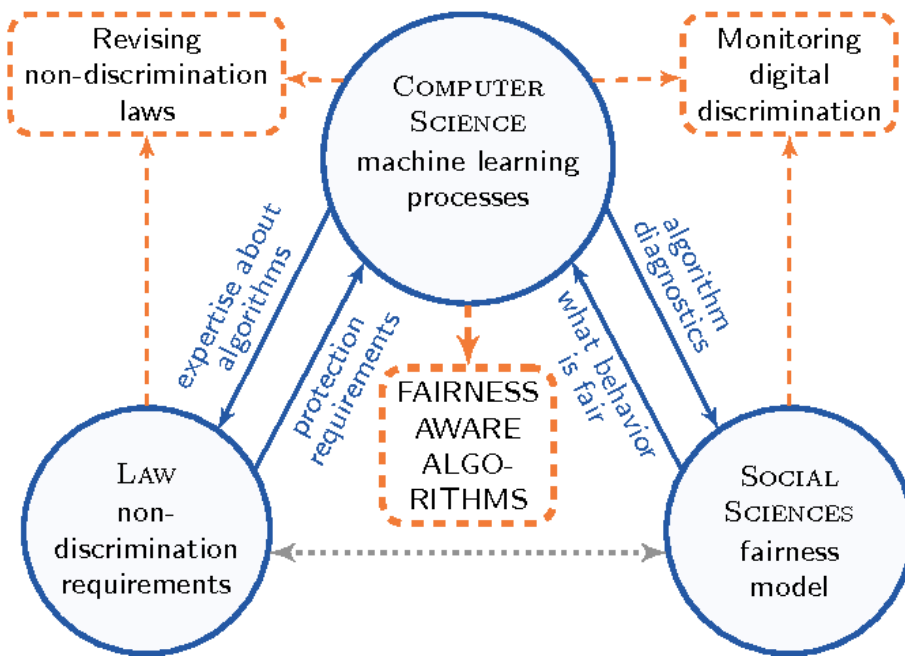
Η πρώτη συνθήκη είναι αναγκαία αλλά όχι επαρκής για την διασφάλιση της εξάλειψης των διακρίσεων, καθώς μολονότι άτομα του πληθυσμού με παρόμοια χαρακτηριστικά λαμβάνουν την ίδια μεταχείριση, υπάρχει περιθώριο για διαφορετική μεταχείριση σε άτομα από διαφορετικές ομάδες του πληθυσμού. Όπως αναφέρεται παρακάτω στην παράγραφο 4.3, η εκπλήρωση της πρώτης συνθήκης μπορεί να διασφαλίσει την εξάλειψη των άμεσων διακρίσεων, διακρίσεων δηλαδή οι οποίες πραγματοποιούνται ανάμεσα σε δύο διαφορετικά άτομα.

Αντίστοιχα η δεύτερη συνθήκη μπορεί να διασφαλίσει την εξάλειψη των έμμεσων διακρίσεων, διακρίσεων δηλαδή οι οποίες πραγματοποιούνται ανάμεσα σε άτομα από διαφορετικές ομάδες του πληθυσμού.

Χαρακτηριστικά αξίζει να αναφερθεί ότι στο παρελθόν κάτοικοι συγκεκριμένων περιοχών αξιολογούνταν αρνητικά από χρηματοπιστωτικά ιδρύματα για την παροχή δανείων. μολονότι φαίνεται ότι το κριτήριο αυτό δεν εμπεριέχει κάποια διάκριση και παραβίαση των ανθρωπίνων δικαιωμάτων, εντούτοις καθώς σε διαφορετικές γειτονίες κατοικούν ομάδες διαφορετικών φυλών, τα χρηματοπιστωτικά ιδρύματα επέλεξαν εις γνώση τους να αποκλείσουν από το δανεισμό άτομα από διαφορετικές φυλές. Στο σενάριο αυτό αν πραγματοποιούνταν σύγκριση ανάμεσα σε άτομα διαφορετικού φύλου δεν θα ήταν εφικτό να διαγνωστεί η ύπαρξη της διάκρισης λόγω της φυλετικής καταγωγής, η οποία εντούτοις αποτελεί μια έμμεση διάκριση και κατάφορη παραβίαση των ανθρωπίνων δικαιωμάτων.

Επιπρόσθετα, στα πλαίσια προσέγγισης και ορισμού του δικαίου κατά την αλγοριθμική λήψη αποφάσεων θα πρέπει να καταστεί σαφές πόσο εντοπισμός και η αναγνώριση διακρίσεων μπορεί να γίνει μόνον όταν οι μεταβλητές στόχοι – αποφάσεις μπορούν να καταταγούν σε σειρά σημαντικότητας, δηλαδή μόνον όταν μια απόφαση μπορεί να χαρακτηριστεί καλύτερη ή χειρότερη από μια άλλη απόφαση. Χαρακτηριστικό παράδειγμα αποτελεί η έγκριση ή η μη έγκριση παροχής ενός δανείου. η έγκριση είναι μια θετική απόφαση, καλύτερη από την αρνητική απόφαση της μη έγκρισης του δανείου.

Επιπρόσθετα, σε ότι αφορά στα προστατευόμενα χαρακτηριστικά, δεν υπάρχουν ιδιαίτεροι περιορισμοί ως προς τον τύπο μεταβλητών με τις οποίες αυτά εξετάζονται. ως εκ τούτου τα προστατευόμενα χαρακτηριστικά μπορούν να εξετάζονται από κατηγορικές, αριθμητικές ή δυαδικές μεταβλητές. Εντούτοις, το φύλο μπορεί να εξεταστεί με μια δυαδική μεταβλητή, το εισόδημα μπορεί να εκφραστεί με μια αριθμητική μεταβλητή και η φυλετική καταγωγή ή το επίπεδο εκπαίδευσης μπορεί να εξεταστεί με κατηγορικές μεταβλητές. Αξίζει επιπλέον να σημειωθεί ότι η πιθανότητα λήψης μεροληπτικών αλγοριθμικών αποφάσεων δεν μπορεί και δεν πρέπει να εξετάζεται αποκλειστικά σε μοναδικά προστατευόμενα χαρακτηριστικά, καθώς οι αλγοριθμικές αποφάσεις μπορεί να μεροληπτούν για άτομα με συνδυασμό προστατευόμενων χαρακτηριστικών. Για παράδειγμα, ένα χρηματοπιστωτικό ίδρυμα μπορεί να εγκρίνει δάνεια για άγαμους άνδρες που ανήκουν σε κάποια εθνική μειονότητα, αλλά να μην εγκρίνει δάνεια για άγαμες γυναίκες που ανήκουν στην ίδια εθνική μειονότητα.



Εικόνα 7: Αλληλεπίδραση γνώσεων, τεχνικών και εργαλείων για τη δίκαια αλγοριθμική λήψη αποφάσεων (Žliobaitė, 2017b)

Έχοντας ορίσει τις συνθήκες υπό τις οποίες η αλγοριθμική λήψη αποφάσεων πραγματοποιείται με σεβασμό στο δίκαιο και στα ανθρώπινα δικαιώματα πρέπει να αναφερθεί ότι η μέτρηση των διακρίσεων και ο σχεδιασμός των αλγορίθμων και προβλεπτικών μοντέλων λήψης αποφάσεων πρέπει να γίνεται με τρόπο τέτοιο ώστε να αποκλείεται η λήψη οποιασδήποτε μεροληπτικής απόφασης λόγω οποιουδήποτε είδους ή οποιουδήποτε συνδυασμού προστατευόμενων χαρακτηριστικών του εξεταζόμενου πληθυσμού.

Κεφάλαιο 3^ο: Ανθρώπινα Δικαιώματα και Διακρίσεις

Σε αυτό το κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από την ανασκόπηση στο πεδίο των ανθρωπίνων δικαιωμάτων και των διακρίσεων, υπό το πρίσμα της αλγοριθμικής λήψης αποφάσεων. Περισσότερη έμφαση δίδεται στην εφαρμογή της αλγοριθμικής λήψης αποφάσεων στη δικαιοσύνη, με εκτενή ανάλυση του συνόλου των συμβάσεων και νομοθετικών πλαισίων τα οποία διέπουν την απονομή δικαιοσύνης και τα σχετικά ανθρώπινα δικαιώματα.

3.1 Εισαγωγή

Στην Διακήρυξη των Ανθρωπίνων Δικαιωμάτων, το πρώτο ανθρώπινο δικαίωμα αναφέρει ότι όλοι οι άνθρωποι γεννιούνται ίσοι, με αξιοπρέπεια και δικαιώματα. Οι άνθρωποι εκ φύσεως διαθέτουν λογική και συνείδηση και πρέπει να ενεργούν με καλοσύνη και σεβασμό μεταξύ τους.

Το δεύτερο άρθρο αναφέρεται στον καθορισμό των δικαιωμάτων και των υποχρεώσεων ακόμα και για αυτούς που αντιμετωπίζουν τυχόν ποινικές κατηγορίες εναντίον τους και αφορά την ισότητα και την απαγόρευση διακρίσεων. Έτσι, όλοι έχουν το δικαίωμα να ακουστούν πλήρως και ισότιμα από ανεξάρτητο και αμερόληπτο δικαστήριο σε δίκαιη και ανοιχτή ακρόαση. Διακρίσεις που σχετίζονται με την εθνικότητα ή την κοινωνική καταγωγή, την περιουσία, τον τόπο γέννησης ή άλλες συνθήκες, θεωρούνται απαράδεκτες. Όλοι οι άνθρωποι πρέπει να τυγχάνουν ίσης μεταχείρισης, ανεξάρτητα από το αν προέρχονται από χώρα με διεθνή αναγνώριση, από χώρα υπό εντολή του ΟΗΕ ή από χώρα που δεν αποικίστηκε ποτέ.

Η έννοια των ανθρωπίνων δικαιωμάτων περιλαμβάνει και την ανάπτυξη και χρήση νέων τεχνολογιών, η οποία κατά τις τελευταίες δεκαετίες έχει κατακλύσει την ζωή μας. Ολοένα και μεγαλύτερος αριθμός ειδικών, κρούουν τον κώδωνα του κινδύνου όσον αφορά την σημασία των ανθρωπίνων δικαιωμάτων στην ανάλυση του κοινωνικού αντίκτυπου της χρήσης της τεχνολογίας.[12]

Σε αρκετές περιπτώσεις διεθνείς οργανισμοί χρησιμοποιούν τα ανθρώπινα δικαιώματα, ως το μέσον ελέγχου της ανάπτυξης και χρήσης των νέων τεχνολογιών. Δεδομένου ότι η χρήση των νέων τεχνολογιών έχουν εισέλθει στη ζωή όλων, τις

τελευταίες δεκαετίες και η νομοθεσία που αφορά στην προστασία των δικαιωμάτων των χρηστών είναι διαρκώς υπό ανάπτυξη.[12]

3.2 Κίνδυνοι διακρίσεων

Πολλοί ερευνητές χαρακτηρίζουν τα συστήματα τεχνητής νοημοσύνης και τα συστήματα αλγοριθμικής λήψης αποφάσεων συχνά ως μαύρα κουτιά. Πολλές φορές δεν είναι κατανοητό από όλους γιατί ένα σύστημα λαμβάνει μια συγκεκριμένη απόφαση. Λόγω της αδιαφάνειας τέτοιων αποφάσεων, είναι δύσκολο για τους ανθρώπους να αξιολογήσουν εάν υπέστησαν διακρίσεις, για παράδειγμα λόγω της φυλετικής καταγωγής ή άλλων δημογραφικών χαρακτηριστικών (Wischmeyer, 2020).

Η αλγοριθμική λήψη αποφάσεων και δη η λήψη αποφάσεων μέσω τεχνητής νοημοσύνης μπορεί να οδηγήσει σε διακρίσεις με διάφορους τρόπους. Σύμφωνα με τους Selbst και Barocas (2018) οι αποφάσεις που λαμβάνονται αλγοριθμικά είναι συχνά ανεξήγητες, ειδικά από όλους όσοι δεν είναι εξειδικευμένοι στον τομέα της τεχνητής νοημοσύνης. Στη βάση αυτού του συμπεράσματος διέκριναν πέντε τρόπους με τους οποίους η αλγοριθμική λήψη αποφάσεων μπορεί να οδηγήσει ακούσια σε διακρίσεις. Διακρίσεις κατά την αλγοριθμική λήψη αποφάσεων μπορεί να οφείλονται (Selbst & Barocas, 2018):

- στον τρόπο με τον οποίον οποίο ορίζεται η μεταβλητή στόχος – η έκβαση την οποία καλείται να αξιολογήσει ένα σύστημα λήψης αποφάσεων και στα χρησιμοποιούμενα κριτήρια αξιολόγησης

Προκειμένου για τη λήψη αποφάσεων, οι αλγόριθμοι που χρησιμοποιούνται αναζητούν συσχετίσεις ανάμεσα σε σύνολα δεδομένων και μεταβλητών. Για παράδειγμα, όταν δημιουργούνται φίλτρα για τον αυτόματο χαρακτηρισμό ηλεκτρονικών μηνυμάτων ως ανεπιθύμητα, εισάγεται ένα σύνολο χαρακτηριστικών των μηνυμάτων, όπως ο αποστολέας, το domain του αποστολέα, λέξεις ή φράσεις κλειδιά κ.α., τα οποία τα χαρακτηρίζουν ως ανεπιθύμητα. Κάθε ένα από αυτά τα χαρακτηριστικά ονομάζεται μεταβλητή στόχος.

Επιπρόσθετα, προκειμένου για τη γρηγορότερη εκτέλεση των αλγορίθμων, συχνά οι πιθανές τιμές των μεταβλητών στόχων χρησιμοποιούνται για σκοπούς

ομαδοποίησης. Για παράδειγμα, προκειμένου για την αυτόματη πρώτη αξιολόγηση αιτούντων για θέσεις εργασίας πρέπει να απαντηθεί το ερώτημα «τι χαρακτηρίζει έναν καλό υπάλληλο;». Η απάντηση αυτού του ερωτήματος μπορεί να είναι μία μεταβλητή στόχος και με βάση τις τιμές της οι αιτούντες μπορούν να διαχωριστούν σε δύο σύνολα: καλούς και κακούς υποψήφιους εργαζομένους. Το πρόβλημα κατά τον ορισμό της τιμής της μεταβλητής η οποία χαρακτηρίζει έναν καλό εργαζόμενο είναι ότι μπορεί να δημιουργεί διακρίσεις για συγκεκριμένες ομάδες υποψηφίων. Για παράδειγμα, εργαζόμενοι με χαμηλότερο εισόδημα πιθανόν κατοικούν σε προάστια και χρειάζεται να χρησιμοποιήσουν μέσα μαζικής μεταφοράς ως εκ τούτου χρειάζεται περισσότερος χρόνος για να μεταβούν στην εργασία τους. Ως εκ τούτου, αυξάνει η πιθανότητα να καθυστερούν να φτάσουν εγκαίρως στη δουλειά τους λόγω κυκλοφοριακής συμφόρησης σε σχέση με υποψηφίους με μεγαλύτερο εισόδημα οι οποίοι μπορεί να κατοικούν στα κέντρα των πόλεων ή να χρησιμοποιούν το δικό τους αυτοκίνητο. Επιπρόσθετα, η χρήση της συχνότητας αργοπορίας ως μεταβλητή στόχο για τη δημιουργία ομάδων υποψηφίων θα οδηγούσε σε διακρίσεις και έναντι άλλων υποψηφίων, όπως υποψηφίων με ειδικές ανάγκες και κινητικά προβλήματα, μεταναστών, μονογονέων κ.ο.κ.. Ως εκ τούτου, κατά την αλγοριθμική λήψη αποφάσεων μπορεί να γίνονται διακρίσεις εξαιτίας του τρόπου με τον οποίο ορίζονται οι μεταβλητές στόχοι και με τον οποίο ομαδοποιούνται βάσει αυτών των μεταβλητών οι υποψήφιοι (Vetrò, et al., 2021).

- στην μεροληψία των δεδομένων εκπαίδευσης των προβλεπτικών μοντέλων – αλγορίθμων λήψης αποφάσεων

Σε περίπτωση όπου τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των προβλεπτικών μοντέλων είναι μεροληπτικά, τότε το αποτέλεσμα των αλγορίθμων οδηγεί με βεβαιότητα σε διακρίσεις αναπαράγοντας τη μεροληψία των δεδομένων εκπαίδευσης. Ειδικά στην περίπτωση όπου τα δεδομένα εκπαίδευσης έχουν επιλεγεί από ανθρώπους, τότε τα δεδομένα εκπαίδευσης είναι πιθανόν να φέρουν την τυχόν μεροληψία των ανθρώπων. Αν για παράδειγμα χρησιμοποιηθούν ιστορικά δεδομένα έγκρισης αιτήσεων σπουδαστών, τα οποία προέρχονται από αποφάσεις ανθρώπων που μεροληπτούν έναντι των γυναικών ή ανθρώπων άλλων φυλών, τότε ο αλγόριθμος

επιλογής θα μεροληπτεί εναντίον γυναικών και υποψηφίων από άλλες φυλές (Guidotti, et al., 2018). Επιπρόσθετα, σύμφωνα με τους Lum και Isaac (2016) η χρήση μεροληπτικών ιστορικών δεδομένων εγκληματικότητας για την εκπαίδευση αλγορίθμων λήψης αποφάσεων μπορεί να οδηγεί σε αλγορίθμους οι οποίοι διακρίνουν μειονότητες ή ομάδες με όμοια δημογραφικά χαρακτηριστικά, καθώς βάσει των δεδομένων εκπαίδευσης σχετίζονται με μεγάλο αριθμό αδικημάτων (Lum & Isaac, 2016). Ωστόσο, αξίζει να σημειωθεί πως στη βάση αυτή, οι αλγόριθμοι και τα συστήματα λήψης αποφάσεων δεν εισάγουν νέες προκαταλήψεις ή νέες διακρίσεις απλώς αναπαράγουν τις διακρίσεις και προκαταλήψεις που επικρατούσαν πριν από τη χρήση τους (Lee, 2018).

- στην μεροληψία κατά τη δειγματοληψία των δεδομένων εκπαίδευσης των προβλεπτικών μοντέλων – αλγορίθμων λήψης αποφάσεων

Η ίδια η διαδικασία της δειγματοληψίας μπορεί να είναι μεροληπτική. Για παράδειγμα, κατά τη συλλογή ιστορικών δεδομένων από αδικήματα και εγκληματικές πράξεις, προκειμένου να δημιουργηθούν προφίλ με βάση την πιθανότητα τέλεσης νέων αδικημάτων, η αστυνομία μπορεί να οδηγηθεί σε περισσότερους αναλογικά ελέγχους σε άτομα που έχουν φέρει ή φέρουν την ιδιότητα του μετανάστη. Σύμφωνα με τους Lum και Isaac, όσο περισσότερο οι αρχές εστιάζουν την προσοχή τους στα πλαίσια της αστυνόμευσης και των ελέγχων ατόμων από διαφορετικές εθνοτικές ομάδες, τόσο περισσότεροι θα αυξάνεται η διαθεσιμότητα ιστορικών δεδομένων ατόμων που ανήκουν στις ομάδες αυτές (Lum & Isaac, 2016).

Εάν ένα προβλεπτικό μοντέλο βασισμένο σε τεχνητή νοημοσύνη ή άλλες αλγοριθμικές τεχνικές έχει εκπαιδευτεί με μεροληπτικά δεδομένα εκμάθησης, τότε θα μεροληπτεί στις αποφάσεις του έναντι των ομάδων που υπέρ εκπροσωπούνται στο δείγμα των δεδομένων εκπαίδευσης. Με άλλα λόγια, τα προβλεπτικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων αναπαράγουν όποιες προκαταλήψεις εμπεριέχονται στα δεδομένα εκπαίδευσης (Lum & Isaac, 2016).

Αντίστοιχα, ο Ferguson (2017) προτείνει ότι αν η αστυνομία εστιάζει την περιοχή της σε περιοχές όπου διαμένουν πολλοί μετανάστες, ακόμη και αν στην περιοχή αυτή

τα ποσοστά εγκληματικότητας δεν αποκλίνουν από το μέσο όρο, θα οδηγηθεί στη βεβαίωση περισσότερων παραβάσεων και αδικημάτων. Τα ιστορικά δεδομένα στα οποία καταγράφονται οι παραβάσεις αυτές και τα αδικήματα αυτά θα μεροληπτούν έναντι των μεταναστών (Ferguson, 2017). Στη συνέχεια, πυροδοτείται ένας φαύλος κύκλος όπου η αστυνομία οδηγείται σε ακόμη περισσότερους ελέγχους σε αντίστοιχες γειτονιές αν αξιολογεί προβλεπτικά μοντέλα.

Οι Bond et al. (2019) πρότειναν την ανάγκη εκδημοκρατισμού των δεδομένων εκπαίδευσης των προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων προτού αυτά χρησιμοποιηθούν. Πιο συγκεκριμένα, κατέγραψαν ότι οι κίνδυνοι διακρίσεων και παραβίασης των ανθρωπίνων δικαιωμάτων αποτυπώνονται ως η τάση των προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων να λαμβάνουν μεροληπτικές αποφάσεις βάσει των κοινωνικών, δημογραφικών και άλλων τυχόν προστατευόμενων χαρακτηριστικών των υποκειμένων της αλγοριθμικής λήψης αποφάσεων. Στην κατεύθυνση αυτή πρότειναν την ανάγκη να εξασφαλίζεται ότι τα ιστορικά δεδομένα τα οποία απαρτίζουν τα δεδομένα εκπαίδευσης αποτελούνται από αντιπροσωπευτικά δείγματα ανθρώπων, προκειμένου να αποφεύγεται η αναπαραγωγή διακρίσεων και παραβίασης των ανθρωπίνων δικαιωμάτων (Bond, et al., 2019).

Οι Tene και Polonetsky (2017) κατέγραψαν την περίπτωση χρήσης ηλεκτρονικών υπηρεσιών για τους πολίτες προκειμένου να αναφέρουν κακοτεχνίες στο δίκτυο πεζοδρόμων και οδών με χρήση smartphones. Μέσα από τα συλλεγόμενα δεδομένα, οι αρμόδιες υπηρεσίες μπορούν να προγραμματίζουν την επιδιόρθωση τέτοιων κακοτεχνιών και βλαβών στο οδικό δίκτυο. Στην περίπτωση όπου η εφαρμογή αυτή χρησιμοποιείται σε γειτονιές όπου κατοικούν μειονότητες ή πολίτες με χαμηλότερα εισοδήματα, τότε οι Tene και Polonetsky (2017) προτείνουν ότι υπάρχει μεγαλύτερη πιθανότητα οι πολίτες να μην έχουν στην κατοχή τους έξυπνες συσκευές, με αποτέλεσμα να αποκλείονται από τη δυνατότητα αναφοράς κακοτεχνιών και βλαβών στο οδικό δίκτυο. Στην περίπτωση αυτή, τα δεδομένα τα οποία συλλέγονται προκειμένου να αξιολογηθούν από έναν αλγόριθμο για τον προγραμματισμό επεμβάσεων στο οδικό δίκτυο είναι από μόνα τους μεροληπτικά και αναμένεται να οδηγούν στον προγραμματισμό περισσότερων επεμβάσεων σε άλλες γειτονιές όπου

δεν κατοικούν μειονότητες ή και όπου οι πολίτες έχουν υψηλότερα εισοδήματα (Tene & Polonetsky, 2017).

- στην επιλογή των ανεξάρτητων μεταβλητών βάσει των οποίων ένα σύστημα λήψης αποφάσεων εκπαιδεύεται

Ένα τέταρτο πρόβλημα σχετίζεται με τα χαρακτηριστικά (κατηγορίες δεδομένων) που επιλέγει ένας οργανισμός για το σύστημα τεχνητής νοημοσύνης του. Εάν ένας οργανισμός θέλει να χρησιμοποιήσει AI για να προβλέψει κάτι αυτόματα, πρέπει να απλοποιήσει τον κόσμο για να είναι σε θέση να το καταγράψει σε δεδομένα. 44 Όπως σημειώνουν οι Barocas και Selbst, ένας οργανισμός πρέπει «να επιλέγει ποια χαρακτηριστικά παρατηρεί και στη συνέχεια να αναδιπλώνεται στις αναλύσεις του». 45

Ας υποθέσουμε ότι ένας οργανισμός θέλει να προβλέψει αυτόματα ποιοι αιτούντες εργασία θα είναι καλοί υπάλληλοι. Δεν είναι δυνατόν, ή τουλάχιστον πολύ δαπανηρό, για ένα στέλεχος τεχνητής νοημοσύνης να αξιολογήσει πλήρως κάθε υποψήφιο εργασία. Ένας οργανισμός θα μπορούσε να επικεντρωθεί, για παράδειγμα, σε ορισμένα χαρακτηριστικά ή χαρακτηριστικά κάθε αιτούντος εργασία.

Επιλέγοντας ορισμένα χαρακτηριστικά, ο οργανισμός ενδέχεται να εισαγάγει προκατάληψη κατά ορισμένων ομάδων. Για παράδειγμα, πολλοί εργοδότες στις ΗΠΑ αναζητούν άτομα που σπούδασαν σε διάσημα και ακριβά πανεπιστήμια. Αλλά μπορεί να είναι σχετικά σπάνιο για ορισμένες φυλετικές ομάδες να σπουδάσουν σε αυτά τα ακριβά πανεπιστήμια. Ως εκ τούτου, μπορεί να έχει ως αποτέλεσμα διακρίσεις εάν ένας εργοδότης επιλέγει υποψηφίους για εργασία με βάση το αν σπούδασαν σε διάσημο πανεπιστήμιο. 46 Εν ολίγοις, οι οργανισμοί μπορούν να προκαλέσουν δυσμενείς διακρίσεις επιλέγοντας τα χαρακτηριστικά που χρησιμοποιεί ένα σύστημα ΤΝ για την πρόβλεψη.

- στην ύπαρξη έμμεσων διακρίσεων

Οι Barocas και Selbst (2016) πρότειναν ότι υπάρχει σημαντικός κίνδυνος διάκρισης ή και παραβίασης των ανθρωπίνων δικαιωμάτων αν χρησιμοποιούνται

δεδομένα τα οποία μπορεί να συνδέονται έμμεσα με προστατευόμενα δημογραφικά κοινωνικά ή άλλα χαρακτηριστικά των υποκειμένων της αλγοριθμικής λήψης αποφάσεων χωρίς αυτή η σχέση να λαμβάνεται υπόψη κατά την εκπαίδευση των προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων ή και κατά την θέσπιση των κανόνων λήψης αποφάσεων (Barocas & Selbst, 2016). Ένα χαρακτηριστικό παράδειγμα τέτοιας έμμεσης διάκρισης μπορεί να εντοπιστεί στο χρηματοπιστωτικό σύστημα. Οι Lepri et al. (2018) πρότειναν ότι κατά την αξιολόγηση του πιστωτικού κινδύνου αιτούντων δανείων είναι πιθανό οι τράπεζες να προβαίνουν σε έμμεσες διακρίσεις κατά τη χρήση των αλγορίθμων και μοντέλων υπολογισμού του πιστωτικού κινδύνου. Πιο συγκεκριμένα, σε περίπτωση όπου ο ταχυδρομικός κώδικας των αιτούντων χρησιμοποιείται ως ανεξάρτητη μεταβλητή για τον υπολογισμό του πιστωτικού κινδύνου, τότε εισάγεται έμμεση διάκριση σε βάρος των αιτούντων οι οποίοι έχουν χαμηλότερα εισοδήματα και κατοικούν σε περιοχές με χαμηλότερες αντικειμενικές αξίες. Αντίστοιχα, αιτούντες που ανήκουν σε μειονότητες ή οι οποίοι έχουν διαφορετική εθνοτική καταγωγή μπορεί ομοίως να κατοικούν σε συγκεκριμένες περιοχές. Ως εκ τούτου, μολονότι ένα χρηματοπιστωτικό ίδρυμα το οποίο χρησιμοποιεί τον ταχυδρομικό κώδικα ως ανεξάρτητη μεταβλητή για τον υπολογισμό του πιστωτικού κινδύνου δεν μεροληπτεί εναντίον αιτούντων με χαμηλότερα εισοδήματα, από μειονότητες, ή από άλλες εθνοτικές καταγωγές, εντούτοις λειτουργούν έμμεσες διακρίσεις σε βάρος τους, καθώς η περιοχή στην οποία κατοικούν συνδέεται με τα παραπάνω κοινωνικά και δημογραφικά, προστατευόμενα χαρακτηριστικά (Lepri, et al., 2018).

Οι Barocas και Selbst (2016) χαρακτηρίζουν τη χρήση δεδομένων κάτω από τα οποία μπορεί να υποφώσκουν ιδιαίτερα κοινωνικά, δημογραφικά και άλλα προστατευόμενα χαρακτηριστικά ως περιττή κωδικοποίηση. Στην πραγματικότητα, ο ταχυδρομικός κώδικας είναι εξαιρετικά απίθανο να χρησιμεύει ως ανεξάρτητη μεταβλητή για τον υπολογισμό του πιστωτικού κινδύνου. Ειδικά δεδομένου ότι μπορεί να συνδέεται με προστατευόμενα χαρακτηριστικά των υποκειμένων της αλγοριθμικής λήψης αποφάσεων, και δεδομένης της χαμηλής αξίας του ως ανεξάρτητη μεταβλητή για τον υπολογισμό του πιστωτικού κινδύνου, πρέπει να μην

συμπεριλαμβάνεται κατά τον σχεδιασμό των κανόνων λήψης απόφασης για αντίστοιχες εφαρμογές (Barocas & Selbst, 2016).

Ο Duguay (2016) πρότεινε ότι στα μέσα κοινωνικής δικτύωσης υπάρχει πρόσφορο έδαφος για έμμεσες διακρίσεις. Πιο συγκεκριμένα, το σύνολο των φίλων και ακολούθων των χρηστών στα μέσα κοινωνικής δικτύωσης χρησιμοποιείται κατά κόρον για τον σεξουαλικό προσανατολισμό των χρηστών χωρίς αυτή να το γνωρίζουν. Η έμμεση διάκριση απορρέει από το γεγονός ότι χρήστες οι οποίοι έχουν φίλους ή ακολούθους συγκεκριμένου σεξουαλικού προσανατολισμού ταυτοποιούνται από τα μέσα κοινωνικής δικτύωσης και τους χρησιμοποιούμενους αλγόριθμους σαν να έχουν και αυτοί τον ίδιο σεξουαλικό προσανατολισμό (Duguay, 2016).

Όπως προαναφέρεται, στην κατεύθυνση εξάλειψης των έμμεσων διακρίσεων προτείνεται η αποφυγή των περιττών κωδικοποιήσεων, δηλαδή η μη χρησιμοποίηση μεταβλητών οι οποίες είναι πιθανόν να συνδέονται με δημογραφικά, κοινωνικά και άλλα τυχόν προστατευόμενα χαρακτηριστικά των υποκειμένων της αλγοριθμικής λήψης αποφάσεων. Ωστόσο, δεδομένου ότι το σύνολο των μεταβλητών οι οποίες μπορεί να υποκρύπτουν προστατευόμενα χαρακτηριστικά είναι αχανές, η λύση αυτή θα αφαιρούσε μεγάλο κομμάτι από τις δυνατότητες πρόβλεψης και λήψης απόφασης από τους αλγόριθμους. Ως εκ τούτου, οι Barocas και Selbst (2016) πρότειναν τη μείωση τις υπολογιζόμενης ακριβείας των αποφάσεων που λαμβάνονται από προοπτικά μοντέλα και αλγόριθμους, προκειμένου η ισχύς των αποφάσεων αυτή να είναι μικρότερη, ειδικά στην περίπτωση όπου υπάρχει κίνδυνος απόκρυψης προστατευόμενων χαρακτηριστικών πίσω από τις ανεξάρτητες μεταβλητές των προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων (Barocas & Selbst, 2016).

Επιπρόσθετα, σε άλλο άρθρο των Barocas και Selbst (2016) προτείνεται μία ακόμη αιτία για τη δημιουργία διακρίσεων κατά την αλγοριθμική λήψη αποφάσεων· τα προβλεπτικά μοντέλα και συστήματα τεχνητής νοημοσύνης μπορούν να κατασκευάζονται κατά βούληση και ως εκ τούτου μπορεί να χρησιμοποιούνται σκοπίμως για την εισαγωγή διακρίσεων (Barocas & Selbst, 2016). Πιο συγκεκριμένα,

σε περίπτωση όπου αυτός ο οποίος σχεδιάζει ένα προβλεπτικό μοντέλο ή έναν αλγόριθμο λήψης αποφάσεων είναι ο ίδιος προκατειλημμένος σε βάρος συγκεκριμένων ανθρώπων, τότε είναι εξαιρετικά πιθανόν η προκατάληψη αυτή να εμποδωθεί και στον αλγόριθμο λήψης αποφάσεων. Αντίστοιχα, δεδομένου ότι τα προβλεπτικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων σχεδιάζονται για λογαριασμό εταιρειών, οργανισμών και φορέων, σε περίπτωση όπου οι διοικήσεις αυτών είναι προκατειλημμένες σε βάρος συγκεκριμένων ομάδων του πληθυσμού, τότε είναι εξαιρετικά πιθανόν οι αλγόριθμοι λήψης αποφάσεων και τα προβλεπτικά μοντέλα τα οποία σχεδιάζονται για λογαριασμό αυτών των οντοτήτων να είναι σκοπίμως μεροληπτικά σε βάρος των ομάδων του πληθυσμού αυτών (Lepri, et al., 2018).

Οι Tene και Polonetsky (2017), εξέτασαν την περίπτωση αμερικάνικης αλυσίδας καταστημάτων λιανικής, η οποία χρησιμοποιούσε αλγόριθμο για τον υπολογισμό της πιθανότητας οι πελάτισσές της να κυοφορούν. Ο αλγόριθμος αυτός αποσκοπούσε στην παροχή εξατομικευμένων προσφορών και διαφημίσεων σε γυναίκες οι οποίες κυοφορούν. Ο υπολογισμός πιθανότητας εγκυμοσύνης πραγματοποιούνταν μέσα από την ανάλυση των καταναλωτικών συνηθειών και συμπεριφορών των γυναικών πελατών της αλυσίδας. Μολονότι ο αλγόριθμος αυτός δεν φαντάζει να προβαίνει σε διακρίσεις σε βάρος των γυναικών, εντούτοις παραβιάζει κατάφορα την ιδιωτικότητά τους χωρίς αυτές να το γνωρίζουν και ενώ απλώς εξακολουθούν να ψωνίζουν σε συγκεκριμένη αλυσίδα λιανικής (Tene & Polonetsky, 2017).

3.3 Η χρήση αλγοριθμικής λήψης αποφάσεων στην Δικαιοσύνη

Η τεχνολογία υπηρετεί εδώ και καιρό την δικαιοσύνη όπως και άλλους κλάδους. Η επικράτηση της τεχνολογίας ξεκίνησε με χρήση ενισχυτικού εξοπλισμού όπως εκτυπωτές, φωτοαντιγραφικά μηχανήματα και προσωπικούς υπολογιστές. Επιπλέον, χρησιμοποιούνται πλήθος ψηφιακών εφαρμογών, που αφορούν συστήματα διαχείρισης, διαχείριση υποθέσεων, ηλεκτρονική κατάθεση και νομικές βάσεις δεδομένων. Για παράδειγμα, στην Εσθονία το 2005 αναπτύχθηκε και εφαρμόστηκε το σύστημα e – file. Στο σύστημα αυτό δημιουργήθηκε μια ηλεκτρονική πλατφόρμα,

στην οποία ο πολίτης έχει την δυνατότητα να εισέρχεται και να καταχωρεί την υπόθεσή του. Η πλατφόρμα διαθέτει τις απαραίτητες δικλίδες ασφαλείας ώστε τα δεδομένα που εισάγει ο πολίτης σχετικά με την υπόθεση να προστατεύονται με τον καλύτερο δυνατό τρόπο[20]. Οι τεχνολογικές εξελίξεις είναι δυνατόν να ταξινομηθούν ως εξής:

A) συστήματα που αφορούν την ενημέρωση των διαδικών σχετικά με τη νομική διαδικασία, τις επιλογές και γενικά τα συστήματα που είναι διαθέσιμα προκειμένου να στηρίζεται επαρκώς η δικανική διαδικασία.

Η διενέργεια ψηφιακών νομικών διαδικασιών, η ηλεκτρονική επικοινωνία μεταξύ διαδικών, δικαστηρίων και αρχών, η ηλεκτρονική διαβίβαση εγγράφων και η διενέργεια οπτικοακουστικών ακροάσεων και συνεδρίων αποτελούν σημαντικό στοιχείο της αποτελεσματικής απονομής της δικαιοσύνης σε πολλά κράτη μέλη ενώ χρησιμοποιήθηκαν εντατικά λόγω της πανδημίας COVID-19. Στο μέλλον, τα ρομπότ μπορεί να είναι σε θέση να εκτελούν όλο και πιο περίπλοκες εργασίες. Το πρόγραμμα υπολογιστή απέδειξε την ανωτερότητά του σε σχέση με τους ανθρώπινους δικηγόρους και κατέλαβε την πρωτιά σε έναν διαγωνισμό μεταξύ των δύο ομάδων [8].

Η ευρεία χρήση της πληροφορικής από τους δικηγόρους οδήγησε στη δημιουργία νέων βάσεων δεδομένων χρησιμοποιώντας προσαρμοσμένο λογισμικό για τις ανάγκες κάθε πελάτη. Η ιδέα της χρήσης ρομπότ δικηγόρων έχει αρχίσει να καταλαμβάνει έδαφος καθώς χρησιμοποιείται ευρέως από τους δικηγόρους. Η δικηγορική εταιρεία Baker & Hostetler ανακοίνωσε ότι στρατολογεί ρομπότ με το όνομα «Ross» της εταιρείας IBM για να αναλάβει την στελέχωση του τμήματος, το οποίο αποτελείται από μόλις 50 δικηγόρους. Το συγκεκριμένο ρομπότ καλείται να χρησιμοποιήσει την τεράστια υπολογιστική ισχύ του λογισμικού του IBM Watson για να μπει σε τεράστιες βάσεις δεδομένων και έγκαιρα να μάθει πώς να εξυπηρετεί καλύτερα τους πελάτες του. Πρόκληση αποτελεί ο προγραμματισμός του ρομπότ ώστε να απαντά σε κανονικές ερωτήσεις, όχι με κωδικοποιημένες λέξεις,

λαμβάνοντας υπόψη το γεγονός ότι οι δικαστικές αποφάσεις είναι γενικά με το μέρος του κατηγορουμένου [9].

Οι εφαρμογές μηχανικής μάθησης γίνονται όλο και πιο κοινές στα δικαστικά συστήματα πολλών χωρών. Η ευρεία χρήση αλγορίθμων αποτελεί πλέον απαραίτητο εργαλείο για τη λήψη αποφάσεων για ορισμένες πολιτείες των ΗΠΑ από δικαστές. Το λογισμικό COMPAS 25 που αναπτύχθηκε από μια ιδιωτική εταιρεία προορίζεται για την αξιολόγηση του κινδύνου καταδίκης εκ νέου ενός ατόμου που έχει κατηγορηθεί και στο παρελθόν. Διαπιστώθηκε ότι οι Αφροαμερικανοί είχαν διπλάσιο ποσοστό υποτροπής σε σύγκριση με άλλες ομάδες πληθυσμού, αποδεικνύοντας πόσο προκατειλημμένος μπορεί να καταλήξει ο αλγόριθμος, ήδη από το σχεδιασμό του. Συγκεκριμένα, το σύστημα εφαρμόστηκε σε συγκεκριμένη περίπτωση ιδιαίτερης σημασίας στην οποία ο κατηγορούμενος καταδικάστηκε για υπόθεση που αφορούσε πυροβολισμούς από όχημα. Ο κατηγορούμενος ομολόγησε την ενοχή του για μικρότερα αδικήματα [10]. Σε αρκετές περιπτώσεις το δικαστήριο κατά τον υπολογισμό της ποινής έλαβε υπόψη τον αλγόριθμο υπολογισμού κινδύνου Compass, ενώ στη διαμόρφωση του αποτελέσματος σημαντικός παράγοντας αποδείχθηκε το φύλο του κατηγορούμενου. Έτσι, ο κατηγορούμενος θεωρήθηκε άτομο με αυξημένες πιθανότητες υποτροπής.

Ο κατηγορούμενος ζήτησε πρόσβαση ώστε να έχει ίδια γνώση των παραμέτρων υπολογισμού κινδύνου. Μέσω δικαστικής απόφασης, αυτή η πρόσβαση δεν δόθηκε στον κατηγορούμενο, παρόλο που υπέβαλε ένσταση για αντισυνταγματικότητα και παραβίαση του δικαιώματός του σε δίκαιη δίκη. Το δικαστήριο αποφάσισε ότι οι κατηγορούμενοι υπέστησαν άδικη μεταχείριση λόγω του φύλου τους. Η υπόθεση έφτασε στο ανώτατο δικαστήριο του Wisconsin, το οποίο έκρινε ότι η χρήση αυτού δεν παραβίαζε κανένα από τα δικαιώματά του, παρά το γεγονός ότι η μεθοδολογία COMPAS δεν είχε κοινοποιηθεί δεόντως από την εταιρεία που επικαλείται το εμπορικό απόρρητο. Η εταιρεία αρνήθηκε ακόμη και μετά τις έντονες αντιδράσεις τον έλεγχο του αλγορίθμου της.

Σε αντίθεση με τις ΗΠΑ, τα νομικά συστήματα των ευρωπαϊκών χωρών επιλέγουν μια πιο ορθολογική χρήση της τεχνητής νοημοσύνης ενώ ιδιαίτερη σημασία αποδίδεται στη συμμόρφωση των αλγορίθμων με τον Γενικό Κανονισμό Προστασίας Δεδομένων και την Οδηγία για την επιβολή του νόμου. Αυτοί είναι κανονισμοί που διατυπώθηκαν ρητά σχετικά με τα όρια για το απόρρητο που αφορά την συλλογή καθώς και την επεξεργασία δεδομένων. Η χρήση της τεχνητής νοημοσύνης μπορεί να ποικίλλει σημαντικά ανάλογα με τις εφαρμογές που περιλαμβάνουν προηγμένες μηχανές αναζήτησης, νομολογία, ηλεκτρονική επίλυση διαφορών, συνταξιοδοτική βοήθεια, σχέδια αποφάσεων, ανάλυση δεδομένων προγνωστική ανάλυση, κλίμακες κ.λ.π.[11].

3.4 Ευρωπαϊκή Σύμβαση για τα Δικαιώματα του Ανθρώπου

Τα ανθρώπινα δικαιώματα ορίζονται τόσο σε επίπεδο διεθνούς, όσο και σε επίπεδο εσωτερικού δικαίου. Το διεθνές δίκαιο των ανθρωπίνων δικαιωμάτων απαιτεί από τα εθνικά κράτη να σέβονται, να προστατεύουν και να εκπληρώνουν τα ανθρώπινα δικαιώματα. Μέγιστο πανανθρώπινο δικαίωμα αποτελεί η αρχή σύμφωνα με την οποία *«όλα τα ανθρώπινα όντα γεννιούνται ελεύθεροι και ίσοι στην αξιοπρέπεια και τα δικαιώματα»*.[12]

Τα ανθρώπινα δικαιώματα έχουν καθολική ισχύ. Είναι αδιαίρετα, πράγμα που σημαίνει ότι όλα τα ανθρώπινα δικαιώματα έχουν ίσο καθεστώς. Είναι αλληλεξαρτώμενα και αλληλένδετα, υπό την έννοια ότι η ενίσχυση ενός ανθρώπινου δικαιώματος είναι πιθανόν να λειτουργεί ενισχυτικά και στην εφαρμογή άλλων δικαιωμάτων και από την αντίθετη πλευρά, η στέρηση ενός δικαιώματος πιθανά λειτουργεί απαγορευτικά και για άλλα. Τα ανθρώπινα δικαιώματα ενσωματώνουν όλες τις συνιστώσες δικαίου, την ανάπτυξη κατάλληλων πολιτικών για την προώθηση και την εφαρμογή τους και τη λήψη αποφάσεων.[12]

Η Ευρωπαϊκή Σύμβαση για τα Δικαιώματα του Ανθρώπου (ΕΣΔΑ) υπογράφηκε στη Ρώμη στις 4 Νοεμβρίου του 1950, υπό την αιγίδα του Συμβουλίου της Ευρώπης ενώ είναι γνωστή και ως Σύμβαση για την Προστασία των Δικαιωμάτων του Ανθρώπου και των Θεμελιωδών Ελευθεριών. Σκοπός της ΕΣΔΑ

είναι η προάσπιση των ανθρωπίνων δικαιωμάτων και η διαφύλαξη θεμελιωδών ελευθεριών. Η Σύμβαση υπογράφεται από όλα τα κράτη μέλη της Ε.Ε. ενώ όταν εισέρχονται νέα μέλη στην ένωση, πρέπει να την αποδέχονται και να την επικυρώνουν σε σύντομο χρονικό διάστημα.[13]

Η ίδρυση του Ευρωπαϊκού Δικαστηρίου Ανθρωπίνων Δικαιωμάτων στηρίχθηκε στην ΕΣΔΑ. Έτσι, ο πολίτης χώρας – μέλους της Ε.Ε. που θεωρεί ότι έχουν παραβιαστεί τα δικαιώματά του έχει το δικαίωμα να προσφύγει στο εν λόγω ευρωπαϊκό δικαστήριο. Οι αποφάσεις του Ευρωπαϊκού Δικαστηρίου είναι βαρύνουσας σημασίας και έχουν δεσμευτικό χαρακτήρα, ως εκ τούτου σε περίπτωση δικαίωσης του πολίτη, αυτός έχει δικαίωμα αποζημίωσης. Το ευρωπαϊκό καθεστώς προστασίας του πολίτη από παραβιάσεις ανθρωπίνων δικαιωμάτων αποτελεί καινοτομία όσον αφορά τις διεθνείς συμβάσεις ανθρωπίνων δικαιωμάτων, καθώς παρέχει στο άτομο έναν ενεργό ρόλο στη διεθνή σκηνή, ενώ μέχρι το 1950 μόνο τα κράτη θεωρούνταν υποκείμενα της διεθνούς νομοθεσίας. Η ΕΣΔΑ αποτελεί μια διεθνή συμφωνία που παρέχει στο άτομο υψηλότερο βαθμό ασφάλειας για την προστασία του. Επιπλέον, πέραν των ευρωπαϊών πολιτών δικαίωμα παραπομπής σε δίκη έχουν τα κράτη – μέλη μετά από καταγγελία άλλου κράτους – μέλους.[13]

Η ΕΣΔΑ περιέχει πλήθος πρωτοκόλλων. Το Πρωτόκολλο νούμερο 6 απαγορεύει την θανατική ποινή με εξαίρεση τις περιόδους πολέμου, όπου επιβάλλεται στρατιωτικός νόμος στον οποίο επιτρέπεται η θανατική ποινή. Κάθε κράτος – μέλος είναι δυνατόν να αποδεχθεί διαφορετικό πλήθος πρωτοκόλλων, αλλά υπό την επιταγή της πραγματικής προστασίας της ανθρώπινης ύπαρξης επιβάλλεται η αποδοχή κατά το δυνατόν περισσότερων.[13]

Παρόλο που η προστασία των ανθρωπίνων δικαιωμάτων αποτελεί ύψιστο στόχο κάθε ευνομούμενης κοινωνίας, καθίσταται σαφές ότι η ολοένα και περισσότερη χρήση αλγορίθμων εγκυμονεί κινδύνους καταπάτησης τους. Οι κίνδυνοι αυτοί εστιάζουν σε ανισότητες σχετικά με το φύλο, με συγκεκριμένες κοινωνικές ομάδες καθώς και με την αναπαραγωγή και διαίωνη προκαταλήψεων. Αυτό είναι δυνατόν να συμβεί σε περιπτώσεις που στους αλγορίθμους δεν έχουν ενσωματωθεί

όλες οι παράμετροι σχετικά με το ζήτημα [22]. Ως αποτέλεσμα τέτοιων παραλείψεων, τα συστήματα τεχνητής νοημοσύνης οδηγούνται στην εξαγωγή μεροληπτικών συμπερασμάτων σε βάρος κοινωνικών ομάδων με συγκεκριμένα χαρακτηριστικά που αφορούν το φύλο, την καταγωγή, την ταυτότητα, τα θρησκευτικά πιστεύω ή τα πολιτικά φρονήματα.

Σχετικές έρευνες αποδεικνύουν ότι σημαντικές ανισότητες κατά την χρήση αλγοριθμικών συστημάτων προκύπτουν όσον αφορά την καταγωγή, όπου κατά πλειοψηφία θίγονται οι Αφροαμερικανοί και Λατοαμερικανούς. Χαρακτηριστικό παράδειγμα αποτελεί η χρήση της εφαρμογής αναγνώρισης προσώπου (FaceID) που χρησιμοποιείται ολοένα και περισσότερο με σκοπό οι χρήστες να τυγχάνουν μέγιστης ασφάλειας. Με την εν λόγω εφαρμογή οι Αφροαμερικανοί, λανθασμένα, συχνά κατηγοριοποιούνταν ως πιθανοί εγκληματίες.[23]

Τέλος, καθώς τα τελευταία χρόνια γίνεται αποδεκτή η διαφοροποίηση ανάμεσα στο βιολογικό φύλο και την ταυτότητα φύλου, διακρίσεις προκύπτουν από την χρήση της εν λόγω τεχνολογίας και στην περίπτωση των διεμφυλικών ατόμων. Οι αλγόριθμοι έχουν την δυνατότητα να διακρίνουν το φύλο ως προς το βιολογικό μέρος αυτού, δηλαδή όσον αφορά το θηλυκό ή το αρσενικό, αναγνώριση που προκύπτει από την φυσική εμφάνιση και τα χαρακτηριστικά του προσώπου. Σε πολλές περιπτώσεις αυτού του είδους η αδυναμία στο αλγοριθμικό σύστημα δημιουργεί σοβαρά ζητήματα. Χαρακτηριστικό παράδειγμα αποτελεί η δημοσιοποίηση μιας περίπτωσης οδηγού ταξί, που μέσω της συγκεκριμένης εφαρμογής δεν μπορούσε να αποδείξει την επαγγελματική του ιδιότητα με αποτέλεσμα να τεθεί σε διαθεσιμότητα μέχρι την αποκατάσταση του προβλήματος.[23]

Τέλος, καθώς τα αλγοριθμικά συστήματα κατακλύζουν εφαρμογές επιλογής προσωπικού, παρατηρείται άνιση κατανομή ισχύος εφόσον για τις υψηλόβαθμες θέσεις των εταιρειών επιλέγονται και προτείνονται άνδρες. Από τα παραπάνω καθίσταται σαφές ότι η χρήση αλγοριθμικών συστημάτων απαιτεί περαιτέρω ανάπτυξη προκειμένου η εφαρμογή της να συνάδει με την προάσπιση των Ανθρωπίνων Δικαιωμάτων, την Αρχή της Ισονομίας και της ίσης μεταχείρισης.

3.5 Ευρωπαϊκό Δικαστήριο Δικαιωμάτων του Ανθρώπου

Το Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων, ΕΔΑΔ, το οποίο ανεπίσημα αναφέρεται και ως δικαστήριο του Στρασβούργου, εφόσον εδρεύει στη συγκεκριμένη περιοχή, δημιουργήθηκε προκειμένου να βοηθήσει στην προστασία των ανθρωπίνων δικαιωμάτων στην Ευρώπη και εγκρίθηκε από το Συμβούλιο της Ευρώπης το 1950. Το Δικαστήριο είναι επιφορτισμένο με την προστασία των ανθρωπίνων δικαιωμάτων και πρόκειται για τον πρώτο διεθνή δικαστικό μηχανισμό με τον σκοπό αυτό. Από τον Οργανισμό παρέχεται η δυνατότητα τόσο διεθνικής όσο και ατομικής προσφυγής, μετά την εξάντληση των εγχώριων ένδικων μέσων. Το Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων επιβλέπει την πιστή εφαρμογή και επιβολή των αποφάσεων του Ευρωπαϊκού Δικαστηρίου από τα κράτη - μέλη. Κάθε κράτος μέλος της Σύμβασης αντιπροσωπεύεται από έναν δικαστή σε αυτό. Το 2021 ο Ιωάννης Κτιστάκης ξεκινά την εννέα χρόνων θητεία του αντιπροσωπεύοντας την Ελλάδα. Το Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων αποτελεί φορέα απονομής δικαιοσύνης διαφορετικό από το Δικαστήριο της Ευρωπαϊκής Ένωσης και βρίσκεται στο Στρασβούργο της Γαλλίας ενώ το Δικαστήριο της Ευρωπαϊκής Ένωσης εδρεύει στο Λουξεμβούργο και αποσκοπεί στην εφαρμογή του ευρωπαϊκού δικαίου.

Κάθε πολίτης έχει τη δυνατότητα να προσφύγει στο Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων, το οποίο είναι όργανο των Ηνωμένων Εθνών. Δεν επιτρέπεται η χρήση του από δημόσια υπηρεσία ή από δημόσιο φορέα δημοσίου δικαίου. Η διαδικασία αναφοράς δεν είναι μια ατομική διαδικασία. Για την άσκηση της αγωγής αρκεί επιστολή στη Γραμματεία της Διεύθυνσης του δικαστηρίου. Για την άσκηση προσφυγής πρέπει να πληρούνται δύο προϋποθέσεις:

(α) κράτος μέλος της ΕΣΔΑ έχει παραβιάσει οποιαδήποτε από τις διατάξεις της και

(β) ο πολίτης έχει εξαντλήσει τα εσωτερικά (κρατικά) ένδικα μέσα για αυτή την παράβαση. Επίσημος εκπρόσωπος του ελληνικού κράτους στο Δικαστήριο είναι το

Νομικό Συμβούλιο του Κράτους και οι υποθέσεις διεκπεραιώνονται από στελέχη του υπό τις οδηγίες και τις κατευθύνσεις του Υπουργείου Εξωτερικών.

Τα κράτη είναι ελεύθερα να λαμβάνουν αποφάσεις σε περίπτωση όμως που διαπιστωθεί παράβαση, η οποία ακόμη και αν ανατραπεί η απόφαση, δεν αίρεται, το Δικαστήριο έχει την εξουσία να επιδικάσει χρηματική αμοιβή στον πολίτη για κάλυψη των δικαστικών εξόδων και για αποκατάσταση της ηθικής βλάβης που το άτομο υπέστη. Εάν η πράξη παραβίασης ήταν νομοθετικό μέτρο, το Δικαστήριο δεν έχει εξουσία να την ακυρώσει, παρά μόνο να την κρίνει αντίθετη με την Ευρωπαϊκή Σύμβαση Ανθρωπίνων Δικαιωμάτων και να επιδικάσει αποζημίωση σε όσους θίγονται από αυτήν. Η αρνητική δημοσιότητα που προκάλεσε και η βεβαιότητα ότι αν κάποιος αθωωθεί θα αθωωθούν και άλλοι στο μέλλον αρκούν συνήθως για να τροποποιήσει το κράτος τη σχετική νομοθεσία και να προβεί σε διακοπή της όποιας παράβασης.

Στο Ευρωπαϊκό Δικαστήριο Ανθρωπίνων Δικαιωμάτων είναι δυνατόν να προσφύγουν άνθρωποι, οι οποίοι θίγονται από την εφαρμογή συστημάτων αλγορίθμων, δηλαδή συστημάτων τεχνητής νοημοσύνης. Η χρήση της εν λόγω τεχνολογίας, οφείλει να λαμβάνει υπόψιν πλήθος δικαιωμάτων που διαφοροποιούνται από τον τομέα τεχνολογίας και την χρήση του. Η χρήση των εν λόγω συστημάτων πρέπει να περιλαμβάνει σεβασμό για την ανθρώπινη αξιοπρέπεια, το δικαίωμα κοινωνικής ασφάλισης και κοινωνικής συνδρομής, το δικαίωμα προστασίας της ιδιωτικής ζωής, την προστασία των προσωπικών και ευαίσθητων δεδομένων, την εξασφάλιση στην πρόσβαση στη δικαιοσύνη και διεξαγωγή δίκαιης δίκης, την προστασία των καταναλωτών, σημείο καίριας σημασίας για τις επιχειρήσεις, το δικαίωμα ορθής διοίκησης όσον αφορά τον δημόσιο τομέα. Οποιοδήποτε ανθρώπινο δικαίωμα θίγεται από την εφαρμογή συστημάτων τεχνητής νοημοσύνης δίνει το δικαίωμα προσφυγής, ακόμα και στο Ευρωπαϊκό Δικαστήριο Ανθρωπίνων δικαιωμάτων.

3.6 Εθνική Επιτροπή για τα Δικαιώματα του ανθρώπου

Η Εθνική Επιτροπή για τα Δικαιώματα του Ανθρώπου (ΕΕΔΑ) ιδρύθηκε με το [Ν.2667/1998](#) και πρόκειται για ανεξάρτητο συμβουλευτικό όργανο της Πολιτείας κατ' εφαρμογή των [Αρχών των Παρισίων](#), που υιοθέτησε η [Γενική Συνέλευση του ΟΗΕ](#) με την απόφασή της 48/134 «Εθνικοί Θεσμοί για την προώθηση και την προστασία των Ανθρωπίνων Δικαιωμάτων» (ΕΘΑΔ) της 20ης Δεκεμβρίου 1993. Η ΕΕΔΑ αποτελεί τον εθνικό θεσμό προστασίας και προώθησης των δικαιωμάτων του ανθρώπου στη χώρα. Η λειτουργία της Εθνικής Επιτροπής καθορίζεται από τις διατάξεις του νόμου [Ν.4780/2021](#), με τον οποίο μεταβλήθηκαν κάποιες από τις ιδρυτικές διατάξεις. Στο [Ν.4780/2021](#) προβλέπεται η απόκτηση νομικής υπόστασης, λειτουργικής ανεξαρτησίας, διοικητικής και οικονομικής αυτοτέλειας.

Οι λόγοι που ενίσχυσαν την ιδέα δημιουργίας της ΕΕΔΑ αφορούν την αδιάκοπη παρακολούθηση των εξελίξεων στο πεδίο που αφορούν το καθεστώς των ανθρωπίνων δικαιωμάτων, τόσο στην Ελλάδα όσο και στο εξωτερικό, η παροχή συμβουλών στους ειδήμονες προκειμένου να επιτυγχάνεται η χάραξη ορθής κεντρικής πολιτικής για το ζήτημα, από το οποίο ελλοχεύουν πλήθος κινδύνων, καθώς η παραβίαση των ανθρωπίνων δικαιωμάτων επισύρει πλήθος ποινών. Για το λόγο αυτό η Ε.Ε.Δ.Α. επισημαίνει διαρκώς στα αρμόδια πολιτειακά όργανα τα δικαιώματα όλων όσων διαβιούν εντός της ελληνικής επικράτειας και όχι κατά αποκλειστικότητα τα δικαιώματα των Ελλήνων πολιτών. Η αρμοδιότητά της δεν περιορίζεται στην γνωμοδότηση, την σύσταση, τις προτάσεις ή την καταγραφή έκθεσης προκειμένου η χώρα να συντάσσεται με τα διεθνή, περιφερειακά ή και ευρωπαϊκά πρότυπα σχετικά με τη διαφύλαξη των ανθρωπίνων δικαιωμάτων. καθώς επίσης και στην ενθάρρυνση για την επικύρωση διεθνών συνθηκών για τα δικαιώματα του ανθρώπου.

Η ανεξαρτησία της ΕΕΔΑ εξασφαλίζεται μέσω της διττής εντολής που αφορά την προώθηση και την ταυτόχρονη προστασία των ανθρωπίνων δικαιωμάτων. Το νομοθετικό πλαίσιο στο οποίο εντάσσεται και διασφαλίζεται η λειτουργία της περιλαμβάνει ορισμένες με σαφήνεια νομοθετικές εντολές. Η νομοθεσία της ΕΕΔΑ προσδιορίζει με σαφήνεια την αποστολή και το πεδίο αρμοδιότητάς της, ενώ

παράλληλα ερμηνεύει με έναν ιδιαίτερα ευρύ και εποικοδομητικό τρόπο την εντολή, ώστε να προωθείται ένας πιο προοδευτικός ορισμός των δικαιωμάτων του ανθρώπου που να περιλαμβάνει το σύνολο αυτών, όπως αυτά ορίζονται σε διεθνή, περιφερειακά και εθνικά κείμενα και ερμηνεύονται από τα αρμόδια διεθνή δικαιοδοτικά όργανα. Με αυτόν τον τρόπο διασφαλίζεται η οικουμενικότητα, η αλληλεξάρτηση και το αδιαίρετο των ανθρωπίνων δικαιωμάτων.

Επιπλέον, η σύνθεση της επιτροπής περιλαμβάνει ανεξάρτητες Αρχές, πανεπιστημιακούς και ακαδημαϊκούς φορείς από σχολές νομικής, πολιτικών επιστημών και κοινωνιολογίας, εκπροσώπους συνδικαλιστικών οργανώσεων, Μη Κυβερνητικές Οργανώσεις, εκπροσώπους πολιτικών κομμάτων και υπουργείων ώστε να εξασφαλίζεται ένα πολύ υψηλό επίπεδο εμπειρογνωμοσύνης και οι απόψεις και πράξεις της Επιτροπής να καλύπτουν τις διαρκώς μεταβαλλόμενες ανάγκες που απαιτούνται.

Η ΕΕΔΑ έχει ένα σαφές θεσμικό πλαίσιο λειτουργίας και διαδραματίζει πρωταγωνιστικό ρόλο στον συμβιβασμό και την ενημέρωση της Κοινωνίας από την μία πλευρά και της Πολιτείας από την άλλη. Επίσης, διασφαλίζει την διαφύλαξη των ανθρωπίνων δικαιωμάτων τόσο σε θεωρητικό επίπεδο όσο και στην πράξη. Σύμφωνα με το νόμο της, η κύρια αποστολή της ΕΕΔΑ συνοψίζεται στα εξής:

1. Αδιάκοπη παρακολούθηση θεμάτων που άπτονται της προστασίας των ανθρωπίνων δικαιωμάτων, της δημόσιας ενημέρωσης καθώς και της ενίσχυσης της σχετικής θεωρητικής έρευνας στο πεδίο.
2. Ανταλλαγή απόψεων και εμπειριών σε διεθνές επίπεδο με αρμόδια όργανα άλλων κρατών - μελών της Ευρωπαϊκής Ένωσης και διεθνείς οργανισμούς, όπως το [Συμβούλιο της Ευρώπης](#), ο [Οργανισμός για την Ασφάλεια και τη Συνεργασία στην Ευρώπη \(ΟΑΣΕ\)](#) και ο [Οργανισμός Ηνωμένων Εθνών \(ΟΗΕ\)](#)
3. Η διαμόρφωση προτάσεων πολιτικής σχετικής με το ζήτημα (Εθνική Επιτροπή για τα Δικαιώματα του Ανθρώπου, 2022).

3.7 Νομοθετικό Πλαίσιο

Η ουσία των Ανθρώπινων Δικαιωμάτων συνοψίζεται στο Άρθρο 1 της Οικουμενικής Διακήρυξης των Ανθρωπίνων Δικαιωμάτων του Ο.Η.Ε:

«Όλοι οι άνθρωποι γεννιούνται ελεύθεροι και ίσοι στην αξιοπρέπεια και τα δικαιώματα. Είναι προικισμένοι με λογική και συνείδηση, και οφείλουν να συμπεριφέρονται μεταξύ τους με πνεύμα αδελφοσύνης»

Τα ανθρώπινα δικαιώματα από ατομική, νομική, πολιτική, οικονομική και κοινωνική σκοπιά αποτελούν βασική προϋπόθεση για τη φυσική και κοινωνική επιβίωση του ανθρώπου ενώ εκφράζουν την επιθυμία του να ζει ως βιολογικό, πνευματικό και κοινωνικό ον. Τα ανθρώπινα δικαιώματα προκύπτουν από τους άγραφους νόμους (φυσικός νόμος) σε συνδυασμό με το θετικό δίκαιο (γραπτοί νόμοι) που είναι υποχρεωτικοί. Η υπακοή στους νόμους είναι υποχρεωτική και ταυτόχρονα θέμα συνείδησης. Τα ανθρώπινα δικαιώματα είναι ο ακρογωνιαίος λίθος του μέλλοντος που επιδιώκει να οικοδομήσει η ανθρωπότητα. Λειτουργώντας ως διεθνείς κανόνες ή σταθερές, διασφαλίζουν την προστασία όλων των ανθρώπων, όπου κι αν βρίσκονται, από κάθε μορφή πολιτικής, οικονομικής, πολιτιστικής, κοινωνικής και νομικής εκμετάλλευσης ή κατάχρησης ενώ ταυτόχρονα διασφαλίζουν την ίση μεταχείριση και την αποφυγή οποιασδήποτε διάκρισης.

Η θεμελίωση των ανθρωπίνων δικαιωμάτων είναι σημαντική για όλους τους ανθρώπους και έχουν αντίκτυπο στην καθημερινή ζωή. Τα ανθρώπινα δικαιώματα ανήκουν σε όλους και διασφαλίζουν ανθρώπινη, αξιοπρεπή και δίκαιη διαβίωση εντός της οικίας, στο χώρο εργασίας, στην εκπαίδευση, στην καθημερινότητα. Η νομική θεμελίωση των ανθρωπίνων δικαιωμάτων θέτουν τις προϋποθέσεις υπό τις οποίες διασφαλίζεται η αξιοπρεπής διαβίωση και απορρίπτεται η εξευτελιστική για την τιμή και την υπόληψη και μεροληπτική μεταχείριση.

Διεθνή όργανα τα οποία έχουν επιφορτιστεί με την προάσπιση των ανθρωπίνων δικαιωμάτων είναι:

A) Το Γενικό Συμβούλιο των Ηνωμένων Εθνών, Τρίτη Επιτροπή (ΓΣΗΕ): Η Γενική Συνέλευση των Ηνωμένων Εθνών είναι υπεύθυνη για την έγκριση διεθνών συμβάσεων που αφορούν τα ανθρώπινα δικαιώματα και τα σχετικά ζητήματα. Τα εν λόγω ζητήματα αποτελούν αναπόσπαστο μέρος της ημερήσιας διάταξης της Γενικής Συνέλευσης των Ηνωμένων Εθνών. Η Τρίτη Επιτροπή της ΓΣΗΕ για τους Πρόσφυγες, η οποία αποτελείται από το σύνολο των κρατών - μελών, ασχολείται με θέματα σχετικά με ανθρώπινα δικαιώματα, τόσο στην ανθρωπιστική όσο και στην κοινωνική τους διάσταση.

B) Το Συμβούλιο των Ανθρωπίνων Δικαιωμάτων (ΣΑΔ): Αποτελεί ένα βοηθητικό όργανο της Γενικής Συνέλευσης και αποτελείται από 47 χώρες που ανήκουν σε πέντε διαφορετικές γεωγραφικές περιοχές. Οι αρμοδιότητες του ΣΑΔ περιλαμβάνουν, την προώθηση, προστασία των ανθρωπίνων δικαιωμάτων, τον έλεγχο των καταγγελιών σχετικά με παραβιάσεις, την υποβολή προτάσεων για την αντιμετώπιση των παραβιάσεων, την διεξαγωγή Οικουμενικής Περιοδικής Αξιολόγησης (Universal Periodic Review, UPR) των χωρών που είναι κράτη – μέλη και τέλος, προκειμένου να εξελιχθεί περαιτέρω το ζήτημα των Ανθρωπίνων Δικαιωμάτων, στις αρμοδιότητες του εντάσσεται και οι εισηγήσεις που υποβάλλονται στη Γενική Συνέλευση του Ο.Η.Ε. Επιπρόσθετα, στις αρμοδιότητες του Συμβουλίου εντάσσεται η προώθηση του διαλόγου και της συνεργασίας μεταξύ φορών, κυβερνήσεων, περιφερειακών οργανισμών, της κοινωνίας των πολιτών και εθνικών επιτροπών με σκοπό την προάσπιση των δικαιωμάτων του ανθρώπου, την αποφυγή των παραβιάσεων και την αντιμετώπιση καταστάσεων που προκύπτουν από τυχόν ανθρωπιστικές κρίσεις και λαμβάνουν τον χαρακτηρισμό του κατ' επείγοντος.

Γ) Όργανα Συμβάσεων Προστασίας των Ανθρωπίνων Δικαιωμάτων, που υποβάλλουν τις χώρες – μέλη που έχουν υπογράψει τις Συμβάσεις του Ο.Η.Ε. για τα ανθρώπινα δικαιώματα σε έναν περιοδικό έλεγχο των Επιτροπών που έχουν επιφορτιστεί με το ζήτημα, στις επιτροπές αυτές εντάσσονται η Επιτροπή υπό το Διεθνές Σύμφωνο για

τα Ατομικά και Πολιτικά Δικαιώματα, το Διεθνές Σύμφωνο για τα Οικονομικά, Κοινωνικά και Πολιτιστικά Δικαιώματα, η Σύμβαση για την Εξάλειψη όλων των Μορφών Φυλετικών Διακρίσεων, η Σύμβαση κατά των Βασανιστηρίων και άλλων Μορφών Σκληρής, Απάνθρωπης ή Εξευτελιστικής Μεταχείρισης ή Τιμωρίας κ.ά. Τα κράτη υποβάλλουν σχετικές εκθέσεις, οι οποίες υπόκειται σε προφορικό έλεγχο, οπότε και δίδονται οι κατάλληλες συστάσεις με σκοπό την συμμόρφωση των κρατών με τις διεθνείς συμβάσεις.

Δ) Η Ύπατη Αρμοστεία του Ο.Η.Ε. για τους πρόσφυγες, η οποία επιφορτίζεται με την προστασία των προσφύγων. Υπό αυτό το πρίσμα πραγματοποιείται έλεγχος ώστε να διασφαλίζεται η εφαρμογή των διεθνών κανόνων δικαίου από τις κυβερνήσεις, ενώ παρέχει και τα βασικά ανθρώπινα αγαθά, πρόσβαση σε καθαρό νερό, τροφή και στέγη. Στη χώρα μας, το γραφείο της Ύπατης Αρμοστείας για τους Πρόσφυγες λειτουργεί από τον Μάρτιο του 1952.

Τα παραπάνω εντάσσονται στο ελληνικό νομοθετικό σύστημα με το Νόμο 4596/2019, όπου περιλαμβάνει την κύρωση των σχετικών διατάξεων όπως προκύπτουν από την Σύμβαση για την Προάσπιση των Δικαιωμάτων του Ανθρώπου και των Θεμελιωδών Ελευθεριών.

3.8 Διάκριση Κατά των Γυναικών

Ως διάκριση κατά των γυναικών, θεωρείται κάθε πράξη που αποκλείει ή θέτει σε εμφανώς μειονεκτική θέση τα πρόσωπα λόγω φύλου, σεξουαλικού προσανατολισμού και ταυτότητας φύλου, καθώς και κάθε εντολή, παρότρυνση ή συστηματική ενθάρρυνση προσώπων να προβαίνουν σε δυσμενή ή άνιση μεταχείριση άλλων λόγω φύλου, σεξουαλικού προσανατολισμού και ταυτότητας φύλου (Νόμος 4604/2019).

- Βία κατά των γυναικών

Οι διακρίσεις κατά των γυναικών περιλαμβάνει την βία, όπου ο όρος «βία κατά των γυναικών» νοείται ως παραβίαση των ανθρωπίνων δικαιωμάτων και μία μορφή

διάκρισης κατά των γυναικών. Οι μορφές βίας περιλαμβάνουν όλες τις πράξεις βίας βασιζόμενης στο φύλο οι οποίες έχουν ως αποτέλεσμα ή ενδέχεται να έχουν ως αποτέλεσμα, σωματική, σεξουαλική, ψυχολογική ή οικονομική βλάβη ή πόνο για τις γυναίκες, συμπεριλαμβανομένων των απειλών τέλεσης τοιούτων πράξεων, τον εξαναγκασμό ή την αυθαίρετη αποστέρηση της ελευθερίας, είτε αυτή συμβαίνει στο δημόσιο ή τον ιδιωτικό βίο. Πηγή: Συμβούλιο της Ευρώπης (2011), Σύμβαση του Συμβουλίου της Ευρώπης για την Πρόληψη και την Καταπολέμηση της Βίας κατά των Γυναικών και της Ενδοοικογενειακής Βίας, Κωνσταντινούπολη, 11.5.2011 (άρθρο 3).

- Βία στην εργασία

Επιθετική συμπεριφορά, σωματική, ψυχολογική ή λεκτική βία λόγω φύλου, σεξουαλικού προσανατολισμού και ταυτότητας φύλου, εντός εργασιακού χώρου, κατά τη διάρκεια και εξ αφορμής της εργασίας. Πηγή: Νόμος 4604/2019: «Πρόωθηση της ουσιαστικής ισότητας των φύλων, πρόληψη και καταπολέμηση της έμφυλης βίας – Ρυθμίσεις για την απονομή Ιθαγένειας – Διατάξεις σχετικές με τις εκλογές στην Τοπική Αυτοδιοίκηση – Λοιπές διατάξεις».

ΒΙΑΣΜΟΣ (RAPE): Είναι ο εξαναγκασμός ατόμου, με σωματική βία ή με απειλή σοβαρού και άμεσου κινδύνου ζωής ή σωματικής ακεραιότητας, σε συνουσία ή σε άλλη ασελγή πράξη ή σε ανοχή της (άρθρο 336 παράγραφος 1 του Ποινικού Κώδικα).

ΔΙΑΣΤΑΣΗ ΤΟΥ ΦΥΛΟΥ (GENDER DIMENSION): Η διάσταση κάθε ζητήματος που σχετίζεται με το φύλο / με τις διαφορές στη ζωή των ανδρών και των γυναικών. Πηγή: 100 λέξεις για την ισότητα. Γλωσσάριο όρων σχετικών με την ισότητα ανδρών και γυναικών, Υπηρεσία Επισήμων Εκδόσεων των Ευρωπαϊκών Κοινοτήτων, 1998.

Όσον αφορά στη φυλετική ανισότητα, μπορεί αρχικά να επισημανθεί ότι όχι μόνο οι αποφάσεις της ΤΝ είναι επιβαρυντικές για τις γυναίκες, αλλά και η ίδια παρουσία των γυναικών και η ενασχόληση τους με τον κλάδο της πληροφορικής παρουσιάζεται μειωμένη, συγκριτικά με προηγούμενες δεκαετίες, που η ύπαρξη και η χρήση των υπολογιστών δεν ήταν ευρέως διαδεδομένη. Η ανισότητα αυτή γίνεται

εμφανής σε όλα τα στάδια της παραγωγικής διαδικασίας. Χαρακτηριστικό είναι το παράδειγμα της εταιρείας Amazon, η οποία είχε δημιουργήσει ένα εργαλείο, το οποίο σάρωνε βιογραφικά υποψηφίων και, από το φιλτράρισμά του κατέληγε στην απόφαση για πρόσληψη. Μετά από αρκετές έρευνες, παρατηρήθηκε ότι το εργαλείο αυτό αυτομάτως υποβάθμιζε τα βιογραφικά εκείνα που περιείχαν τον όρο «κολέγιο θηλέων» (Cellan-Jones, 2017)

Ακόμη, σύμφωνα με μία έρευνα του Αμερικανικού Ινστιτούτου Οικονομικής Έρευνας, η οποία αφορούσε τα έτη 2010-2012, διαπιστώθηκε ότι οι λατινοαμερικανοί, και ειδικά οι γυναίκες, δημιουργοί λογισμικών λάμβαναν ετησίως μισθό μειωμένο κατά 20% σε σχέση με τους άνδρες συναδέλφους τους (Turner, 2016).

Κεφάλαιο 4^ο: Μέτρα και εργαλεία εξάλειψης των διακρίσεων στους αλγορίθμους

Σε αυτό το κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από τη βιβλιογραφική ανασκόπηση στο πεδίο των διακρίσεων και του σεβασμού των ανθρωπίνων δικαιωμάτων κατά την εφαρμογή της αλγοριθμικής λήψης αποφάσεων στη δικαιοσύνη, τόσο υπό το πρίσμα των προτεινόμενων μέτρων για την εξάλειψη των διακρίσεων όσο και υπό το πρίσμα των διαθέσιμων μέσων για τη μέτρηση της απόδοσης των αλγορίθμων σε ό, τι αφορά στην αποφυγή διακρίσεων.

4.1 Εισαγωγή

Στο αυτό το κεφάλαιο παρουσιάζονται και συζητούνται τα ευρήματα από τη βιβλιογραφική ανασκόπηση στο πεδίο των διακρίσεων και του σεβασμού των ανθρωπίνων δικαιωμάτων κατά την εφαρμογή της αλγοριθμικής λήψης αποφάσεων στη δικαιοσύνη, τόσο υπό το πρίσμα των προτεινόμενων μέτρων για την εξάλειψη των διακρίσεων όσο και υπό το πρίσμα των διαθέσιμων μέσων για τη μέτρηση της απόδοσης των αλγορίθμων σε ό, τι αφορά στην αποφυγή διακρίσεων.

4.2 Διαθέσιμα μέτρα αξιολόγησης της απόδοσης των αλγορίθμων ως προς τις διακρίσεις

Δεδομένης της φύσης του προβλήματος της διάκρισης και παραβίασης των ανθρωπίνων δικαιωμάτων κατά την αλγοριθμική λήψη αποφάσεων, έχει ενταθεί το ενδιαφέρον στην κατεύθυνση ανάπτυξης αλγορίθμων για τη λήψη αποφάσεων με επίγνωση των διακρίσεων. Η κατασκευή αυτών των αλγορίθμων αλλά και η αξιολόγηση του συνόλου των αλγορίθμων που εφαρμόζονται για τη λήψη αποφάσεων απαιτεί την υιοθέτηση μέτρων για την αξιολόγηση της απόδοσης των αλγορίθμων ως προς το περιθώριο διακρίσεων (Le Quy, et al., 2022). Τα μέτρα αυτά μπορεί να εκπορεύονται αμιγώς από το πεδίο της αλγοριθμικής λήψης αποφάσεων ή και από άλλους ερευνητικούς τομείς, στην κατεύθυνση διεύρυνσης των δυνατοτήτων εξάλειψης των διακρίσεων κατά την αλγοριθμική λήψη αποφάσεων (Žliobaitė, 2017a). Η ενίσχυση της διαφάνειας κατά την κατασκευή αλγορίθμων για τη λήψη αποφάσεων μπορεί να οδηγήσει στην ενίσχυση της λογοδοσίας και στη σταδιακή

εξάλειψη των διακρίσεων (De Laat, 2018). Ο Tohan (2019) προτείνει ότι η υιοθέτηση μέτρων για την αξιολόγηση των διακρίσεων τόσο λόγω των δεδομένων όσο και λόγω των ανθρώπων που έχουν αναπτύξει τους αλγορίθμους μπορεί να οδηγήσει στην αποτελεσματική ενίσχυση της μεροληψίας των αλγορίθμων (Tolan, 2019).

Σύμφωνα με τον Žliobaitė (2017a), τα μέτρα για την αξιολόγηση των διακρίσεων μπορούν να κατηγοριοποιηθούν σε μέτρα (Žliobaitė, 2017a):

- στατιστικά
- απόλυτα
- υπό όρους
- κατάστασης

Οι τέσσερις αυτές κατηγορίες μέτρων διάκρισης δεν υποκαθιστούν η μία την άλλη καθώς μετρούν διαφορετικές πτυχές των διακρίσεων. Πιο συγκεκριμένα, τα στατιστικά μέτρα μετρούν έμμεσα τόσο την παρουσία όσο και την απουσία διακρίσεων, αλλά δεν μπορούν να χρησιμοποιηθούν για τη μέτρηση της εξάπλωσης των διακρίσεων σε ένα σύνολο δεδομένων ή αποφάσεων. Τα απόλυτα μέτρα και τα μέτρα υπό όρους μετρούν έμμεσα το μέγεθος των διακρίσεων, ενώ τα μέτρα κατάστασης μετρούν άμεσα το εύρος των διακρίσεων των αλγορίθμων. Επιπλέον τα απόλυτα μέτρα χρησιμοποιούνται για την αξιολόγηση των διακρίσεων λαμβάνοντας υπόψη κάποιο χαρακτηριστικό, βάσει του οποίου μπορεί να γίνεται διάκριση, και βάση της της αλγοριθμικής απόφασης. Αξίζει να σημειωθεί ότι κατά την εφαρμογή απόλυτων μέτρων διακρίσεων δεν λαμβάνεται υπόψη το σύνολο των χαρακτηριστικών των ατόμων στα οποία αφορά η αλγοριθμική λήψη αποφάσεων, αντίθετα γίνεται η υπόθεση ότι όλα τα άτομα έχουν τα ίδια χαρακτηριστικά π.χ. εθνικότητα. Επιπρόσθετα, αναφέρεται ότι τα μέτρα υπό όρους μπορούν να χρησιμοποιούνται ως συμπληρωματικά των απόλυτων μέτρων, προκειμένου για τον εντοπισμό και την εξήγηση διακρίσεων οι οποίες δεν οφείλονται σε κάποιο ειδικό χαρακτηριστικό των ατόμων. Σε ό,τι δε αφορά στα μέτρα κατάστασης, αναφέρεται ότι αυτά έχουν δημιουργηθεί για τη μέτρηση των άμεσων ή έμμεσων διακρίσεων οι οποίες εντοπίζονται κατά το στάδιο της ταξινόμησης των δεδομένων εισόδου ενός

μοντέλου λήψης αποφάσεων. Με λίγα λόγια τα μέτρα κατάστασης στοχεύουν στον εντοπισμό διακρίσεων οι οποίες εκπορεύονται από τα ίδια τα δεδομένα εισόδου ενός μοντέλου λήψης αποφάσεων και πιο συγκεκριμένα για τη μέτρηση της εξάπλωσης των διακρίσεων, προσεγγίζοντας το ποσοστό των ατόμων σε ένα σύνολο δεδομένων τα οποία πιθανώς θα υποστούν όποια διάκριση ή παραβίαση των ανθρωπίνων δικαιωμάτων τους (Žliobaitė, 2017a).

4.2.1 Στατιστικά μέτρα

Τα στατιστικά μέτρα ή πιο ορθά οι στατιστικοί έλεγχοι είναι τα μέτρα τα οποία εντοπίζονται να εφαρμόζονται πρωτίτερα από όλα τα υπόλοιπα προτεινόμενα μέτρα διακρίσεων. όπως προαναφέρεται στοχεύουν στον εντοπισμό έμμεσων διακρίσεων στα δεδομένα εισόδου (Köchling & Wehner, 2020). Η εφαρμογή τους συνιστάται αυτό στατιστικό έλεγχο υποθέσεων προκειμένου να εξεταστεί ο βαθμός στον οποίο είναι πιθανόν να πραγματοποιηθεί μια διάκριση. Συνηθέστερα υπό τη μηδενική υπόθεση δεν υπάρχει διαφορετική μεταχείριση του συνόλου των ατόμων και ενός συγκεκριμένου υποσυνόλου των ατόμων το οποίο χαρακτηρίζεται από κάποιο κοινό δημογραφικό ή άλλο χαρακτηριστικό (Vetrò, et al., 2021). Ως στατιστικός έλεγχος ο οποίος πραγματοποιείται είναι αυτός του ελέγχου του βαθμού κατά τον οποίο ή διαφορά ανάμεσα στις αποφάσεις για τα 2 παραπάνω σύνολα είναι τυχαία ή στατιστικά σημαντική. Αν η διαφορά αυτή δεν είναι τυχαία, τότε η μηδενική υπόθεση απορρίπτεται και το αποτέλεσμα του ελέγχου καταδεικνύει ότι υπάρχει διάκριση. Αξίζει να σημειωθεί ότι υπάρχουν διαφορές ανάμεσα στον κλασικό στατιστικό έλεγχο υποθέσεων και στον στατιστικό έλεγχο για τον εντοπισμό διακρίσεων. πιο συγκεκριμένα το επίπεδο εμπιστοσύνης δεν πρέπει να συγχέεται με το μέγεθος της διαφορετικότητας ανάμεσα στο γενικό και το ειδικό σύνολο, η οποία μπορεί να κυμαίνεται από αμελητέα έως πολύ μεγάλη. Επιπρόσθετα, αν η μηδενική υπόθεση απορριφθεί τότε διαγιγνώσκεται η ύπαρξη διάκρισης. Ωστόσο αν η μηδενική υπόθεση δεν απορριφθεί τότε αυτό δεν σημαίνει πως είναι απίθανο να υπάρχει διάκριση. Εντούτοις υπάρχει μεγάλη πιθανότητα το σύνολο δεδομένων να είναι πολύ μικρό για τον επιτυχή έλεγχο της μηδενικής υπόθεσης (Haeri & Zweig, 2020).

Όπως γίνεται εύκολα αντιληπτό οι στατιστικοί έλεγχοι που εφαρμόζονται για τον εντοπισμό διακρίσεων είναι αυτοί οι οποίοι εφαρμόζονται κατά τον κλασικό στατιστικό έλεγχο υποθέσεων όπως το τεστ χ^2 και το t test. Επιπρόσθετα αξίζει να αναφερθεί πως μολονότι τα στατιστικά μέτρα μπορούν να χρησιμοποιηθούν για τον εντοπισμό έμμεσων διακρίσεων, τα ευρήματα από τους αντίστοιχους στατιστικούς ελέγχους δεν μπορούν να χρησιμοποιηθούν για την κατασκευή περιορισμών οι οποίοι θα εισαχθούν στο στάδιο εκπαίδευσης ενός προβλεπτικού μοντέλου ή ενός αλγορίθμου λήψης αποφάσεων. Ωστόσο, τα αποτελέσματα από την εφαρμογή στατιστικών μέτρων μπορούν να χρησιμοποιηθούν για την προσέγγιση μέτρων διακρίσεων και περιορισμών για τη βελτιστοποίηση της αλγοριθμικής λήψης αποφάσεων. Στην κατεύθυνση αυτή μπορούν να χρησιμοποιούνται όλα τα στατιστικά μεγέθη τα οποία υπολογίζονται μέσω της ανάλυσης των δεδομένων σε ένα σύνολο δεδομένων, όπως η μέση τιμή η τυπική απόκλιση, η διασπορά κ.α..

4.2.2 Απόλυτα μέτρα

Όπως προαναφέρεται τα απόλυτα μετά εφαρμόζονται για την μέτρηση του μεγέθους των διακρίσεων διαμέσου της μέτρησης του μεγέθους της διαφορετικότητας ανάμεσα σε δύο σύνολα ατόμων. Προκειμένου για τον ορισμό των δύο αυτών συνόλων χρησιμοποιείται ένα προστατευόμενο δημογραφικό ή άλλο χαρακτηριστικό, όπως η εθνικότητα ή το φύλο. Αξίζει να σημειωθεί ότι τα απόλυτα μέτρα μπορούν να χρησιμοποιηθούν για την αξιολόγηση διακρίσεων ανάμεσα σε διαφορετικά προστατευόμενα σύνολα ατόμων. Στην περίπτωση αυτή, κάθε προστατευόμενο σύνολο συγκρίνεται με τον γενικό πληθυσμό (π.χ. σύνολα ατόμων διαφορετικών εθνικοτήτων) (Criado & Such, 2019).

Κατά την εφαρμογή των απόλυτων μέτρων αξιολογούνται μεγέθη όπως η μέση διαφορά ανάμεσα σε διαφορετικά σύνολα ατόμων, η σταθμισμένη μέση διαφορά ανάμεσα σε διαφορετικά σύνολα ατόμων και η κανονικοποιημένη διαφορά ανάμεσα σε διαφορετικά σύνολα ατόμων. Επιπρόσθετα, βρίσκει εφαρμογή το μέτρο της περιοχής κάτω από την καμπύλη (Area Under Curve), κυρίως για πολύ μεγάλα σύνολα δεδομένων. Κατά την εφαρμογή του στατιστικού αυτού μέτρου, τα άτομα με διαφορετικά δημογραφικά ή άλλα χαρακτηριστικά χωρίζονται σε διαφορετικές τάξεις

και υπολογίζεται το μέγεθος της πιθανής διάκρισης έναντι μιας από όλες τις τάξεις. Επιπρόσθετα προτείνεται η εφαρμογή του λόγου επίδρασης, ο οποίος ισούται με τον λόγο των ευνοϊκών αποφάσεων για ένα προστατευόμενο σύνολο προς το σύνολο των αποφάσεων. Σημειώνεται ότι ο λόγος επίδρασης είναι το αντίστροφο του λόγου πιθανότητας (Manuel, et al., 2012). Αναφέρεται ενδεικτικά ότι στο δικαστικό σύστημα των ΗΠΑ ορίζεται διάκριση όταν ο λόγος επίδρασης είναι χαμηλότερος από το 80%, αφήνοντας ένα περιθώριο 20% μη ευνοϊκών αποφάσεων σε βάρος προστατευόμενων συνόλων του πληθυσμού ως αποδεκτό. Επιπρόσθετα, προτείνεται η εφαρμογή του λόγου πιθανοτήτων, ο οποίος ισούται με το λόγο του συνόλου των αποφάσεων προς τις ευνοϊκές αποφάσεις (Feldman, et al., 2015).

Σε ότι αφορά στον έλεγχο διακρίσεων για περισσότερα του ενός προστατευόμενα σύνολα ατόμων, αναφέρεται ότι τα απόλυτα μέτρα έχουν σχεδιαστεί ως επί το πλείστο για τη σύγκριση ενός προστατευόμενου συνόλου έναντι του συνολικού πληθυσμού. Προκειμένου για την εφαρμογή απόλυτων μέτρων για τη σύγκριση περισσότερων προστατευόμενων συνόλων μπορεί να γίνεται σύγκριση των αποφάσεων ανάμεσα σε ζεύγη προστατευόμενων συνόλων, ή σύγκριση των αποφάσεων για ένα σύνολο σε σχέση με κάθε ένα από τα υπόλοιπα προστατευόμενα σύνολα, ή σύγκριση του προστατευόμενου συνόλου με την επικρατέστερη ομάδα του πληθυσμού (Calders, et al., 2013).

4.2.3 Μέτρα υπό όρους

Όπως προαναφέρεται, με τα μέτρα υπό όρους αξιολογείται το μέγεθος των διακρίσεων έμμεσα. Πιο συγκεκριμένα, ενώ στα απόλυτα μέτρα λαμβάνεται υπόψη μια μεταβλητή στόχος και μια προστατευόμενη μεταβλητή και γίνεται η υπόθεση ότι κάθε διαφορά ανάμεσα στο προστατευόμενο και στο επικρατέστερο σύνολο του πληθυσμού οφείλεται σε διάκριση, στα μέτρα υπό όρους εξετάζεται ο βαθμός στον οποίον οι διαφορές ανάμεσα στο προστατευόμενο και στο επικρατέστερο σύνολο του πληθυσμού μπορούν να εξηγηθούν από άλλα χαρακτηριστικά ή παράγοντες. Το υπολειπόμενο μέγεθος των διαφορών οφείλεται σε διακρίσεις (Criado & Such, 2019). Για παράδειγμα, ένα ποσοστό της ανεργίας μπορεί να οφείλεται σε διαφορές στο επίπεδο εκπαίδευσης, με το υπολειπόμενο μόνον ποσοστό να οφείλεται στη

διαφορετική εθνικότητα των υποψηφίων. Προκειμένου για τον υπολογισμό των μέτρων υπό όρους υπολογίζεται η μέση διαφορά ανάμεσα σε δύο σύνολα και από αυτήν αφαιρείται το μέρος της μέσης διαφοράς το οποίο μπορεί να εξηγηθεί από γνωστές μεταβλητές. Τα μέτρα τάσης και είναι τα πιο κοινά μέτρα υπό όρους. Πιο συγκεκριμένα, τα μέτρα τάσης αποτυπώνουν την πιθανότητα ένα σύνολο του πληθυσμού να αντιμετωπίζεται με μια συγκεκριμένη απόφαση, π.χ. απόφοιτοι πρωτοβάθμιας εκπαίδευσης να είναι άνεργοι για περισσότερο από 6 μήνες. Αν είναι γνωστά όλα τα μέτρα τάσης τα οποία επενεργούν σε ένα προστατευόμενο σύνολο του πληθυσμού τότε μπορεί να υπολογιστεί η διαφορά η οποία δεν οφείλεται σε διακρίσεις (d'Alessandro, et al., 2017).

4.2.4 Μέτρα κατάστασης

Όπως προαναφέρεται, τα μέτρα κατάστασης χρησιμοποιούνται για τη μέτρηση του εύρους των άμεσων διακρίσεων. Τα μέτρα αυτά βασίζονται στην αναγνώριση από τα ίδια τα υποκείμενα της αλγοριθμικής λήψης αποφάσεων, σε ένα σύνολο ατόμων, του αν βιώνουν διάκριση ή όχι. Στη συνέχεια υπολογίζεται το ποσοστό των ατόμων τα οποία έχουν υποστεί διάκριση (Briggs & Hollmén, 2020). Σημειώνεται ότι με τα μέτρα κατάστασης αξιολογούνται όχι μόνον οι αρνητικές αλλά και οι θετικές διακρίσεις. Κατόπιν υπολογισμού μέτρων κατάστασης πραγματοποιείται έλεγχος, μέσω του οποίου συγκρίνετε κάθε άτομο του προστατευόμενου συνόλου με ένα άτομο στο μη προστατευόμενο (επικρατέστερο) σύνολο προκειμένου να εξεταστεί αν η απόφαση του αλγόριθμου θα ήταν διαφορετική (Briggs & Hollmén, 2020).

4.3 Εργαλεία και τεχνικές αλγοριθμικής λήψης αποφάσεων με επίγνωση των διακρίσεων

Προκειμένου για τη δημιουργία αλγορίθμων λήψης αποφάσεων με επίγνωση των διακρίσεων πρέπει πρωτίστως να αφαιρεθούν οι διακρίσεις από τα σύνολα δεδομένων τα οποία χρησιμοποιούνται για την εκπαίδευση των προβλεπτικών μοντέλων και αλγορίθμων. Στα πλαίσια αυτά, πρέπει να λαμβάνονται μέτρα από το επίπεδο της εξόρυξης δεδομένων (Pedreshi, et al., 2008). Η ανάγκη αυτή έχει

οδηγήσει στην ανάπτυξη του κλάδου της εξόρυξης δεδομένων με επίγνωση των διακρίσεων, ενός κλάδου που συνδυάζει γνώσεις, έννοιες και τεχνικές ανάπτυξης αλγορίθμων, τεχνητής νοημοσύνης, επιστήμης δεδομένων, μεγάλων δεδομένων, κοινωνικών επιστημών και δικαίου (Carmichael, et al., 2016; Mittelstadt, et al., 2016).

Στόχος της εξόρυξης δεδομένων με επίγνωση των διακρίσεων είναι ταυτόχρονα η ανακάλυψη των διακρίσεων και η πρόληψή τους μέσα από κατάλληλες τεχνικές εξόρυξης δεδομένων. Στις προηγούμενες παραγράφους παρουσιάστηκαν τα διάφορα διαθέσιμα μέτρα για την ανακάλυψη διακρίσεων σε σύνολα δεδομένων και αλγόριθμους λήψης αποφάσεων. κοινή συνισταμένη των διαφορετικών τεχνικών εξόρυξης δεδομένων με επί μέρους των διακρίσεων αποτελεί ο έλεγχος και ο εντοπισμός κανόνων συσχετίσεις και ταξινόμησης των δεδομένων προκειμένου να απεμπολούνται οι έμφυτες λόγω των ίδιων των δεδομένων διακρίσεις (Hajian & Domingo-Ferrer, 2012; Zhuang, et al., 2018).

Σύμφωνα με τους Kashid et al. (2017), μία από τις απλούστερες μεθόδους για την ανακάλυψη διακρίσεων συνήθως συνίσταται στην κατασκευή μοντέλων γραμμικής παλινδρόμησης, όπου τα προστατευόμενα χαρακτηριστικά (όπως η εθνική καταγωγή, η φυλή, το φύλο, η ηλικία, η θρησκεία κ.α.) εξετάζονται ως ανεξάρτητες μεταβλητές ως προς την επίδραση τους σε εξαρτημένες μεταβλητές στόχους - αποτελέσματα αποφάσεων (προσλήψεις σε μια εργασία, έγκριση εισαγωγή σε ένα πανεπιστήμιο κ.α.) (Kashid, et al., 2017).

Σε αυτά τα μοντέλα παλινδρόμησης, όταν η τιμή significance υποδεικνύει στατιστική σημαντικότητα τότε η επίδραση των προστατευόμενων χαρακτηριστικών σε μεταβλητές στόχους - αποτελέσματα αποφάσεων είναι στατιστικά σημαντική και καταδεικνύει την ύπαρξη κάποιας διάκρισης. Το δε μέγεθος της διάκρισης εξαρτάται από τους υπολογιζόμενους συντελεστές παλινδρόμησης. Όσο μεγαλύτεροι ή μικρότεροι είναι αυτοί οι συντελεστές τόσο μεγαλύτερη είναι και η διάκριση, καθώς πρέπει να σημειωθεί πως οι διακρίσεις δεν είναι μόνον αρνητικές αλλά μπορεί να είναι και θετικές, δηλαδή άτομα με κάποιο προστατευόμενο χαρακτηριστικό να

λαμβάνουν καλύτερη μεταχείριση από το γενικό ή επικρατέστερο μέρος του πληθυσμού (Kashid, et al., 2017).

Έχει αναπτυχθεί πλειάδα αλγορίθμων πρόληψης των διακρίσεων μέσω των οποίων καθίσταται εφικτή η κατασκευή προβλεπτικών μοντέλων και αλγορίθμων για τη λήψη αποφάσεων χωρίς διακρίσεις. Ωστόσο, προκειμένου οι αλγόριθμοι αυτοί να είναι όσο το δυνατόν πιο αποτελεσματικοί στην κατεύθυνση παραγωγής αποφάσεων χωρίς διακρίσεις, πρέπει να είναι διαθέσιμος μεγάλος όγκος δεδομένων, έτσι ώστε να μπορεί να εξεταστεί το μεγαλύτερο δυνατό εύρος προστατευόμενων χαρακτηριστικών τα οποία οδηγούν σε διακρίσεις. Κατόπιν εντοπισμού των αιτιών των διακρίσεων, οι αιτίες αυτές μπορούν να εισάγονται στη φάση κατασκευής των προβληματικών μοντέλων και αλγορίθμων λήψης αποφάσεων ως περιορισμοί. γίνεται εύκολα αντιληπτό ότι οι περιορισμοί αυτοί συνήθως αφορούν στον τρόπο με τον οποίο ένας αλγόριθμος ή ένα προβλεπτικό μοντέλο αξιολογεί και μεταχειρίζεται το σύνολο των δημογραφικών και άλλων χαρακτηριστικών των ατόμων τα οποία υπόκεινται στην αλγοριθμική λήψη αποφάσεων.

Εστιάζοντας περισσότερο στην εξόρυξη δεδομένων με επίγνωση των διακρίσεων μπορεί να αναφερθεί ότι οι τεχνικές για την πρόληψη των διακρίσεων κατατάσσονται σε τρεις μεγάλες κατηγορίες:

- προεπεξεργασία δεδομένων
- μετά επεξεργασία δεδομένων
- κανονικοποίηση προβλεπτικών μοντέλων

Κατά την προεπεξεργασία δεδομένων τα ιστορικά δεδομένα εισόδου τροποποιούνται προκειμένου να αφαιρούνται τυχόν στατιστικά σημαντικές διαφορές ανάμεσα σε άτομα με προστατευόμενα και μη προστατευόμενα χαρακτηριστικά. Κατόπιν τα δεδομένα αυτά χρησιμοποιούνται για την εκπαίδευση των προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων. Με τον τρόπο αυτό αφαιρείται η πηγή της διάκρισης με αποτέλεσμα οι αλγόριθμοι οι οποίοι εκπαιδεύονται με τέτοια σύνολα δεδομένων να είναι βέβαιο ότι δεν οδηγούν στη λήψη αποφάσεων με διακρίσεις. σημειώνεται ότι κατά την προεπεξεργασία των δεδομένων μπορεί να

τροποποιούνται όχι μόνον τα προστατευόμενα χαρακτηριστικά (ανεξάρτητες μεταβλητές) αλλά και οι μεταβλητές στόχοι – αποφάσεις ή και τα δύο ανάλογα με τη στατιστική ισχύ της επίδρασης των ανεξάρτητων στις εξαρτημένες μεταβλητές (Calmon, et al., 2017; du Pin Calmon, et al., 2018).

Κατά την μετά επεξεργασία των δεδομένων, κατασκευάζεται ένα προβλεπτικό μοντέλο και κατόπιν αξιολογείται η απόδοσή του σε ό τι αφορά στη λήψη αποφάσεων χωρίς διακρίσεις. με βάση την αξιολόγηση της απόδοσης του εισάγονται και τροποποιούνται περιορισμοί ή αφαιρούνται κανόνες λήψης απόφασης προκειμένου κατόπιν της ολοκλήρωσης μετά επεξεργασίας των δεδομένων ο αλγόριθμος λήψης αποφάσεων το προβλεπτικό μοντέλο να οδηγεί στη λήψη αποφάσεων χωρίς διακρίσεις (Favaretto, et al., 2019).

Κατά την κανονικοποίηση των προβλεπτικών μοντέλων εισάγονται περιορισμοί για την αφαίρεση των διακρίσεων στη φάση της εκπαίδευσης του μοντέλου. αυτό μπορεί να πραγματοποιηθεί με διάφορους τρόπους όπως η τροποποίηση των κριτηρίων διαχωρισμού και ταξινόμησης αν εφαρμόζονται τεχνικές δένδρων αποφάσεων (Kamiran, et al., 2013).

Γίνονται σαφές ότι και για τις τρεις τεχνικές απαιτείται ο σχεδιασμός, καθορισμός και προσδιορισμός συνεκτικών μέτρων για την αναγνώριση των διακρίσεων. Τα μέτρα αυτά όπως παρουσιάστηκαν στην παράγραφο 4.2 μπορούν να βρίσκουν εφαρμογή είτε κατά μόνος και είτε συνδυαστικά προκειμένου να διευρύνεται το πεδίο εντοπισμού διακρίσεων και να βελτιώνεται η απόδοση των αλγορίθμων λήψης αποφάσεων και προβλεπτικών μοντέλων σε ό τι αφορά στην λήψη αποφάσεων χωρίς διακρίσεις. επιπλέον η επιλογή των μέτρων διακρίσεων μπορεί να γίνεται ανάλογα με το είδος της διάκρισης, δηλαδή ανάλογα με το αν μια διάκριση είναι άμεση ή έμμεση, καθώς όπως προαναφέρεται στο παρόν κεφάλαιο οι διακρίσεις μπορεί να οφείλονται σε αποφάσεις κατά την κατασκευή των προβλεπτικών μοντέλων ή κατά την εκπαίδευση των προβλεπτικών μοντέλων ή κατά την αρχική επεξεργασία και ταξινόμηση των δεδομένων εισόδου και των δεδομένων εκπαίδευσης, αλλά και κατά την επιλογή των περιορισμών οι οποίοι μπορεί να

εισάγονται ή να μην εισάγονται σε έναν αλγόριθμο λήψης αποφάσεων ή σε ένα πραγματικό μοντέλο ακόμη και κατά βούληση προκειμένου για την παραγωγή αποφάσεων με εκούσιες διακρίσεις. πρώτο παρουσιασθούν περισσότερα στοιχεία για την αλγοριθμική λήψη αποφάσεων με επίγνωση των διακρίσεων πρέπει να αναφερθεί ότι οι άμεσες διακρίσεις αναφέρονται σε αποφάσεις που διαφέρουν από άτομο σε άτομο, ενώ οι έμμεσες διακρίσεις αναφέρονται σε αποφάσεις που διαφέρουν ανάμεσα σε διαφορετικές ομάδες του πληθυσμού (Ntoutsis, et al., 2020).

Τα προβληματικά μοντέλα απλώς χρησιμοποιούνται για τη λήψη αποφάσεων εκπαιδεύονται με βάση τα ιστορικά δεδομένα τα οποία παίζουν το ρόλο των δεδομένων εισόδου και τα οποία μπορεί να είναι δεδομένα που αφορούν:

- στο φύλο, στην εθνική καταγωγή, στο θρήσκευμα, στη γλώσσα ομιλίας, στο επίπεδο εκπαίδευσης, στο ποινικό μητρώο κ.α., ειδικά για την αξιολόγηση προ της πρόσληψης σε μία θέση εργασίας ή για την αξιολόγηση της πιθανότητας τέλεσης κάποιου αδικήματος
- στο εισόδημα, στο καθεστώς απασχόλησης, στο πιστωτικό προφίλ, στη φορολογική και ασφαλιστική ενημερότητα κ.α., ειδικά για την πρόβλεψη του πιστωτικού κινδύνου
- στις αναζητήσεις στο διαδίκτυο, στους φίλους στα social media, σε λέξεις κλειδιά τις οποίες καταγράφουν μικρόφωνα έξυπνων συσκευών, σε προτιμήσεις μουσικής, ταινιών κ.α., ειδικά για την εξατομίκευση των διαφημίσεων και προσφορών που λαμβάνουν ή οι οποίες αναπαράγονται στους χρήστες του Διαδικτύου
- σε δεδομένα προηγούμενων αγορών, σε δεδομένα διαδρομών, για την εξατομίκευση προσφορών και εκπτώσεων από ιδιωτικούς οργανισμούς (λιανική, πάρκινγκ, εταιρείες εκμετάλλευσης αυτοκινητοδρόμων)

Με βάση τα δεδομένα εισόδου ένα προβλεπτικό μοντέλο ή ένας αλγόριθμος λήψης αποφάσεων παράγει μια απόφαση ή μια πρόβλεψη. Στα πλαίσια εφαρμογής της αλγοριθμικής λήψης αποφάσεων με επίγνωση των διακρίσεων, όπως αναφέρεται παραπάνω πρέπει να προηγείται ο εντοπισμός η αναγνώριση και η εξάλειψη

διακρίσεων οι οποίες είναι έμφυτες στα δεδομένα εισόδου ή στις αποφάσεις των προγραμματιστών οι οποίοι δομούν τους κανόνες λήψης απόφασης, προκειμένου το αποτέλεσμα της εξόρυξης δεδομένων θα είναι ελεύθερο τόσο από άμεσες όσο και από έμμεσες διακρίσεις. Με τον τρόπο αυτό, και δεδομένου ότι τα προβλεπτικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων βασίζονται σε κανόνες λήψης αποφάσεων, καθίσταται βέβαιο ότι οι προβλέψεις – αποφάσεις θα είναι αμερόληπτες.

Πολλοί ακαδημαϊκοί διαφωνούν σε ό,τι αφορά στη χρήση προστατευόμενων χαρακτηριστικών ως δεδομένων εισόδου. Ενδεικτικά προβληματικά μοντέλα τα οποία αξιολογούν τον πιστωτικό κίνδυνο δεν είναι απαραίτητο να γνωρίζουν το φύλο του υποκειμένου της αξιολόγησης, ούτε και την εθνική καταγωγή το θρήσκευμα ή το επίπεδο εκπαίδευσης του. Με τον τρόπο αυτόν εξασφαλίζεται ότι τα προβληματικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων αυτοί θα παράγουν προβλέψεις – αποφάσεις οι οποίες θα είναι ελεύθερες διακρίσεων χωρίς να δαπανώνται πόροι για την εξόρυξη δεδομένων με επίγνωση των διακρίσεων.

Ωστόσο αξίζει να αναφερθεί ότι ανάλογα με τη φύση των προστατευόμενων δεδομένων τα οποία εξαιρούνται από τα δεδομένα εισόδου στο σύνολο δεδομένων εκπαίδευσης των προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων μπορεί να εξασφαλίζεται η εξάλειψη μόνον των άμεσων ή έμμεσων διακρίσεων και μόνον για τα προστατευόμενα δεδομένα τα οποία έχουν εξαιρεθεί και όχι ταυτόχρονα τόσο των άμεσων και έμμεσων διακρίσεων.

Ειδικά σε ό,τι αφορά στην εξάλειψη των έμμεσων διακρίσεων, από μόνη της η εξαίρεση προστατευόμενων δεδομένων τα οποία μπορεί να χαρακτηρίζουν ομάδες του πληθυσμού, όπως η εθνική καταγωγή και το θρήσκευμα, δεν μπορούν να εξασφαλίσουν την *a priori* εξάλειψη των έμμεσων διακρίσεων. Προκειμένου να εξακριβωθεί ο βαθμός στον οποίο τηρούνται οι απουσίες διακρίσεων και να εξασφαλιστεί η δίκαιη παραγωγή αποφάσεων – προβλέψεων ανάμεσα σε ομάδες του πληθυσμού, θα πρέπει τα προβλεπτικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων να έχουν πρόσβαση στο σύνολο των προστατευόμενων χαρακτηριστικών των ατόμων που απαρτίζουν όλες τις ομάδες του πληθυσμού στα διαθέσιμα ιστορικά δεδομένα.

Αν τα προστατευόμενα δεδομένα δεν είναι διαθέσιμα κατά τη διάρκεια της διαδικασίας εκπαίδευσης των αλγορίθμων λήψης απόφασης και των προβλεπτικών μοντέλων, τότε δεν είναι εφικτό να εφαρμοστούν τα μέτρα διάκρισης τα οποία παρατέθηκαν στην παράγραφο 4.2 και ως εκ τούτου δεν είναι δυνατόν τα προβλεπτικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων να μπορούν να αναγνωρίσουν την ύπαρξη διακρίσεων.

Εντούτοις, τα προβλεπτικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων που κατασκευάζονται χωρίς πρόσβαση σε προστατευόμενα δεδομένα μπορεί να μην κάνουν διακρίσεις εξαιτίας των δεδομένων εισόδου, αλλά στην πραγματικότητα είναι επιρρεπή σε έμμεσες διακρίσεις λόγω του αλγορίθμου και συγκεκριμένα λόγω των κανόνων λήψης απόφασης οι οποίοι έχουν διαμορφωθεί κατά την εκπαίδευση του μοντέλου.

Συμπεράσματα

Σε αυτό το κεφάλαιο εξάγονται συμπεράσματα κατόπιν ολοκλήρωσης της βιβλιογραφικής ανασκόπησης, συζητούνται περιορισμοί των συμπερασμάτων αυτών και δίνονται προτάσεις για μελλοντική έρευνα.

Η εξέλιξη και η διείσδυση της τεχνολογίας στις περισσότερες πτυχές της καθημερινής σύγχρονης ζωής, όπως η ιατρική, η δικαιοσύνη, η ασφάλεια, η εκπαίδευση κ.α. υπόσχεται ποικίλα οφέλη. Στην παρούσα καταγράφονται οφέλη τα οποία σχετίζονται με την μείωση των ανθρώπινων λαθών, με την επιτάχυνση των χρόνων εξυπηρέτησης, τη βελτίωση του επιπέδου εξυπηρέτησης, τη βελτίωση της ακρίβειας των αποφάσεων οι οποίες λαμβάνονται κ.α.. Η εξέλιξη της τεχνολογίας επιτρέπει την κατασκευή προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων με την εισαγωγή κανόνων, νόμων, ηθικών περιορισμών, άρχων κ.α., προκειμένου να εξασφαλίζεται ότι οι αποφάσεις οι οποίες λαμβάνονται είναι αντικειμενικές. Ωστόσο όπως καταγράφεται στην παρούσα, δεδομένου ότι οι αλγόριθμοι λήψης αποφάσεων και τα προβληματικά μοντέλα σχεδιάζονται και κατασκευάζονται από ανθρώπους, είναι εξαιρετικά πιθανόν αυτά να προβαίνουν σε διακρίσεις ή και παραβιάσεις των ανθρωπίνων δικαιωμάτων εκούσια ή ακούσια.

Οι διακρίσεις στις οποίες μπορεί να προβαίνουν τα προβληματικά μοντέλα και οι αλγόριθμοι λήψης αποφάσεων μπορεί να απορρέουν από τα ιστορικά δεδομένα τα οποία χρησιμοποιούνται για την εκπαίδευση τέτοιων μοντέλων. Στην κατεύθυνση αυτή προτείνεται η κανονικοποίηση των ιστορικών δεδομένων προκειμένου να αντιπροσωπεύονται ισάξια όλες οι ομάδες του πληθυσμού σε ένα δείγμα δεδομένων εκπαίδευσης. Επιπρόσθετα, πρέπει να αξιοποιούνται ως ανεξάρτητες μεταβλητές για την πρόβλεψη ή λήψη απόφασης μόνον εκείνες οι μεταβλητές οι οποίες είναι απαραίτητο να χρησιμοποιηθούν και μάλιστα πρέπει να εξετάζεται η πιθανότητα ανεξάρτητες μεταβλητές οι οποίες φαντάζουν αμιγώς αντικειμενικές, να κρύβουν από πίσω τους διακρίσεις σε βάρος συγκεκριμένων ομάδων του πληθυσμού. Επιπλέον, υπάρχει μεγάλος κίνδυνος εκούσιων διακρίσεων οι οποίες μπορεί να απορρέουν από προκαταλήψεις των οργανισμών για τους οποίους σχεδιάζονται προβλεπτικά μοντέλα

και αλγόριθμοι λήψης αποφάσεων ή από τους ίδιους τους ανθρώπους οι οποίοι σχεδιάζουν και κατασκευάζουν προαιρετικά μοντέλα και αλγόριθμους λήψης αποφάσεων. Ειδικά σε ό,τι αφορά σε περιπτώσεις παραβίασης των ανθρωπίνων δικαιωμάτων, οι χρησιμοποιούμενοι αλγόριθμοι λήψης αποφάσεων είναι πιθανόν να εξυπηρετούν τα συμφέροντα των οργανισμών για τους οποίους έχουν κατασκευαστεί προβαίνοντας σε κατάφωρες παραβιάσεις ανθρωπίνων δικαιωμάτων.

Κατά το σχεδιασμό προβλεπτικών μοντέλων και αλγορίθμων λήψης αποφάσεων πρέπει να αξιολογείται η απόδοσή τους σε ότι αφορά στην παραγωγή μεροληπτικών αποφάσεων και διακρίσεων. Στην παρούσα παρουσιάστηκαν και συζητήθηκαν ενδελεχώς μέτρα για την αξιολόγηση της απόδοσης των αλγορίθμων και προβλεπτικών μοντέλων σε ότι αφορά στην παραγωγή διακρίσεων. Αξίζει να σημειωθεί πως οι αλγόριθμοι και η λήψη αποφάσεων με επίγνωση των διακρίσεων φαντάζει ως η πλέον αποδοτική λύση προκειμένου να αποφεύγονται οι διακρίσεις και οι παραβιάσεις των ανθρωπίνων δικαιωμάτων από αλγόριθμους λήψης αποφάσεων και προβλεπτικά μοντέλα.

Σε ό,τι αφορά στο νομοθετικό πλαίσιο, καταγράφονται εντατικές προσπάθειες αυστηροποίησης και εμπλουτισμού του νομοθετικού πλαισίου καθώς διεισδύει περισσότερο η αλγοριθμική λύση αποφάσεων σε ολοένα και περισσότερες πτυχές της σύγχρονης καθημερινής ζωής. Προτείνεται ότι το ισχύον νομοθετικό πλαίσιο είναι πλήρες και προβλέπει το μεγαλύτερο εύρος των πιθανών διακρίσεων και παραβιάσεων των ανθρωπίνων δικαιωμάτων από αλγόριθμους και προβλεπτικά μοντέλα.

Μολονότι κατά το σχεδιασμό αλγορίθμων λήψης αποφάσεων και προβλεπτικών μοντέλων θα πρέπει να λαμβάνεται υπόψη το σύνολο των αρχών, ηθικών περιορισμών, νόμων, κανονισμών κ.α. εντούτοις θα πρέπει να εξασφαλίζεται η βιωσιμότητα των αλγορίθμων λήψης αποφάσεων και προβλεπτικών μοντέλων. Κατόπιν αξιολόγησης των κινδύνων δημιουργίας διακρίσεων και παραβίασης των ανθρωπίνων δικαιωμάτων αλλά και του συνόλου των πλεονεκτημάτων από την εξέλιξη και διείσδυση της αλγοριθμικής λήψης αποφάσεων, το πρόσημο της

τεχνολογίας αυτής είναι θετικό και εναπόκειται στους υπεύθυνους ανάπτυξης τέτοιων αλγορίθμων και προβλεπτικών μοντέλων, αλλά και στις αρχές προκειμένου να εξασφαλίζεται η τακτική αναθεώρηση και ο τακτικός εμπλουτισμός του νομοθετικού πλαισίου, η δημιουργία αλγορίθμων οι οποίοι θα είναι αντικειμενικοί και δεν θα μεροληπτούν σε βάρος κανενός αλλά ούτε και θα προβαίνουν σε παραβιάσεις των ανθρωπίνων δικαιωμάτων.

Βιβλιογραφία

- [1]: Δούμπος, Μ., & Ζοπουνίδης, Κ.: *Λήψη αποφάσεων με πολλαπλά κριτήρια: Μια εισαγωγή στις βασικές έννοιες, μεθοδολογία και εφαρμογές*, Πολυτεχνείο Κρήτης, Τμήμα Μηχανικών Παραγωγής και Διοίκησης
- [2]: Παναγιώτου Ν., (2011): «*Συστήματα Αποφάσεων – Εισαγωγή στην ανάλυση αποφάσεων*» Εθνικό Μετσόβειο Πολυτεχνείο, Τμήμα Μηχανολόγων - Μηχανικών
- [3]: Brian, C., & Griffiths, T.: *Η αλγοριθμική τέχνη των αποφάσεων, Η επιστήμη των υπολογιστών στην καθημερινή ζωή*, Πανεπιστημιακές Εκδόσεις Κρήτης.
- [4]: Νανόπουλος Α., & Μανωλόπουλος Ι. (2008): *Εισαγωγή στην εξόρυξη και τις αποθήκες δεδομένων*. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- [5]: Saaty, T.L. (2008): *Decision making with the analytic hierarchy process*, International Journal of Services Sciences, 1(1), 83-98
- [6]: Ross, J., (2010): *Fuzzy Logic with Engineering Applications*, 3rd edition, Wiley
- [7]: Γκρίτζαλης Δ.Α., Γκρίτζαλης Σ. & Κάτσικας Σ. (2003): *Ασφάλεια δικτύων υπολογιστών*. Παπασωτηρίου
- [8]: The robot lawyers are here - and they're winning, Cellan-Jones, R., 1.11.2017, <https://www.bbc.com/news/technology-41829534>
- [9]: Meet 'Ross,' the newly hired legal robot, Turner, K., Washington Post, 16.05.2016: <https://www.washingtonpost.com/news/innovations/wp/2016/05/16/meet-ross-the-newly-hired-legal-robot/>
- [10]: COMPAS: Correctional Offender Management Profiling for Alternative Sanctions
- [11]: Συμπεράσματα του Συμβουλίου «Πρόσβαση στη δικαιοσύνη – Αξιοποίηση των ευκαιριών της ψηφιοποίησης», Συμβούλιο της Ευρωπαϊκής Ένωσης Βρυξέλλες, 8 Οκτωβρίου 2020, 11599/20, σε: <https://data.consilium.europa.eu/doc/document/ST-11599-2020-INIT/el/pdf> 23

- [12]: Australian Human Rights Commission: Human Rights and Technology, Final Report, 2021
- [13]: Σύμβαση για την Προστασία των Δικαιωμάτων του Ανθρώπου και των Θεμελιωδών Ελευθεριών όπως τροποποιήθηκε από το Πρωτόκολλο υπί αριθ. 11 συνοδευόμενη από τα Πρωτοκόλλα υπ' αριθ. 1, 6, 7, 12 και 13,
- [14]: Wirtz, B.,(2011), “Media and Internet Management”. Ηλεκτρονικό, διαθέσιμο στο http://berndwirtz.com/downloads/mim_lm_extract.pdf
- [15]: Zhang, H., (2010), “Social Media and Distance Education”, Instructional Support Specialist Center for Support of Instruction, δημοσιευμένο στον ιστότοπο: <http://deoracle.org/online-pedagogy/emerging-technologies/social-media-anddistance-education.html/>
- [16]: Kaplan, A. M. & Haenlein, M., (2010), “Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons”, Vol. 53, pp. 59-68.
- [17]: Wellman, B. and Milena, G. (1999), “Net surfers don’t ride alone”, ed. Boulder, Colorado: Westview Press
- [18]: Evans D, (2008), “Social Media Marketing: an hour a day”, Wiley Publishing, Inc. Indianapolis
- [19]: Appgene.net: Αλγόριθμοι αναζητήσεων: Τι είναι και πώς λειτουργούν.
- [20]: F. Contini “Artificial Intelligence and the Transformation of Humans, law and Technology, Interaction in Judicial Proceedings” Law, Technology and Humans, Vol 2 (1) 2020, σελ. 2
- [21]: <https://www.nchr.gr/eeda.html>: ΕΕΔΑ, το ανεξάρτητο συμβουλευτικό όργανο της Πολιτείας
- [22]: Ι στοσελίδα Ευρωπαϊκού Κοινοβουλίου, Τεχνητή νοημοσύνη: ευκαιρίες και απειλές, 26.03.2021, δημοσιευμένο σε: <https://www.europarl.europa.eu/>, [τελευταία επίσκεψη: 04.04.2021]
- [23]: Benno Keller, Promoting responsible artificial intelligence in insurance, 16.01.2020, δημοσιευμένο σε: The Geneva

Association, <https://www.genevaassociation.org/>, [τελευταία επίσκεψη: 07.05.2021]

- [24]: FRA, European Union Agency For Fundamental Rights, Κατανοώντας το Μέλλον, Τεχνητή Νοημοσύνη και Θεμελιώδη Δικαιώματα – Σύνοψη https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_el.pdf.
- [25]: Alikhademi, K. και συν., 2022. A review of predictive policing from the perspective of fairness. *Artificial Intelligence and Law*, 30(1), pp. 1-17.
- [26]: Bader, V. & Kaiser, S., 2019. Algorithmic decision-making? The user interface and its role for human involvement in decisions supported by artificial intelligence. *Organization*, 26(5), pp. 655-672.
- [27]: Barocas, S. & Selbst, A., 2016. Big data's disparate impact. *California law review*, 104(671), pp. 671-732.
- [28]: Bell, M., 2021. Next-generation policing research: three propositions. *Journal of Economic Perspectives*, 35(4), pp. 29-48.
- [29]: Bolukbasi, T. και συν., 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Advances in neural information processing systems*, pp. 4349 - 4357.
- [30]: Bond, R. M. M. και συν., 2019. Human Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making. *Proceedings of RoCHI*, October, pp. 2-9.
- [31]: Briggs, E. & Hollmén, J., 2020. Mitigating Discrimination in Clinical Machine Learning Decision Support Using Algorithmic Processing Techniques. Στο: A. Appice, G. Tsoumakas, Y. Manolopoulos & S. Matwin, επιμ. *Discovery Science: 23rd International Conference, DS 2020*. Thessaloniki: Springer International Publishing, pp. 19-33.
- [32]: Brown, A. και συν., 2019. Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-12.

- [33]: Calders, T. και συν., 2013 . Controlling attribute effect in linear regression. *IEEE 13th international conference on data mining*, December, pp. 71-80.
- [34]: Calders, T. & Zliobaite, I., 2013. Why unbiased computational processes can lead to discriminative decision procedures. Στο: B. Custers, T. Calders, B. Schermer & T. Zarsky, επιμ. *Discrimination and privacy in the information society*. 3 επιμ. Berlin: Springer, p. 43–57.
- [35]: Calmon, F., Wei, D. V. B., Natesan Ramamurthy, K. & Varshney, K., 2017. Optimized pre-processing for discrimination prevention. *Advances in neural information processing systems*, p. 30.
- [36]: Carmichael, L., Stalla-Bourdillon, S. & Staab, S., 2016. Data mining and automated discrimination: a mixed legal/technical perspective. *IEEE Intelligent Systems*, 31(6), pp. 51-55.
- [37]: Cellan-Jones, R., 2017. *The robot lawyers are here - and they're winning*.
 [Ηλεκτρονικό]
 Available at: <https://www.bbc.com/news/technology-41829534>
 [Πρόσβαση 17 Δεκέμβριος 2022].
- [38]: Corbett-Davies, S. και συν., 2017. Algorithmic decision making and the cost of fairness. *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* , pp. 797-806.
- [39]: Covington, P., Adams, J. & Sargin, E., 2016. Deep neural networks for youtube recommendations. *Proceedings of the 10th ACM conference on recommender systems* , pp. 191-198.
- [40]: Criado, N. & Such, J., 2019. Digital discrimination. *Algorithmic regulation*, pp. 82-97.
- [41]: d'Alessandro, B., O'Neil, C. & LaGatta, T., 2017. Conscientious classification: A data scientist's guide to discrimination-aware classification. *Big data*, 5(2), pp. 120-134.
- [42]: De Laat, P., 2018. Algorithmic decision-making based on machine learning from big data: can transparency restore accountability?. *Philosophy & technology*, 31(4), pp. 525-541.

- [43]: Djuric, U., Zadeh, G., Aldape, K. & Diamandis, P., 2017. Precision histology: how deep learning is poised to revitalize histomorphology for personalized cancer care. *npj Precision Oncology*, 1(22), pp. 1-22.
- [44]: Dobbie, W., Goldin, J. & Yang, C. S., 2018. The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review*, 108(2), pp. 201-240.
- [45]: Dobra, A., 2009. Decision Tree Classification. Στο: L. Liu & M. Özsu, επιμ. *Encyclopedia of Database Systems*. Boston, MA: Springer, p. 765–769.
- [46]: du Pin Calmon, F. και συν., 2018. Data pre-processing for discrimination prevention: Information-theoretic optimization and analysis. *IEEE Journal of Selected Topics in Signal Processing*, 12(5), pp. 1106-1109.
- [47]: Duggento, A. και συν., 2021. Deep computational pathology in breast cancer. *Seminars in cancer biology*, Τόμος 72, pp. 226-237.
- [48]: Duguay, S., 2016. He has a way gayer Facebook than I do”: Investigating sexual identity disclosure and context collapse on a social networking site. *New media & society*, 18(6), pp. 891-907.
- [49]: Dwork, C. και συν., 2012. Fairness through awareness. *Proceedings of the 3rd innovations in theoretical computer science conference* , January, pp. 214-226.
- [50]: Euromed Marseille School of Management, World Med MBA Program - Information Systems and Strategy Course, 2017. *Information Systems and Strategy, Session 1, Types of Information System and the Classic Pyramid Model*. [Ηλεκτρονικό]
Available at: http://www.chris-kimble.com/Courses/World_Med_MBA/Types-of-Information-System.html
[Πρόσβαση 28 Νοέμβριος 2022].
- [51]: Favaretto, M., De Clercq, E. & Elger, B., 2019. Big Data and discrimination: perils, promises and solutions. A systematic review. *Journal of Big Data*, 6(1), pp. 1-27.

- [52]: Feldman, M. και συν., 2015 . Certifying and removing disparate impact. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, August, pp. 259-268.
- [53]: Ferguson, A., 2017. Rise of Big Data Policing. *New York University Press*.
- [54]: Fukuchi, K., Kamishima, T. & Sakuma, J., 2015. Prediction with model-based neutrality. *IEICE TRANSACTIONS on Information and Systems*, 98(8), pp. 1503-1516.
- [55]: Furletti, M. J., 2002. An Overview and History of Credit Reporting. *FRB of Philadelphia Payment Cards Center Discussion Paper No. 02-07*, June.
- [56]: Griffard, M., 2019. A Bias-Free Predictive Policing Tool: An Evaluation of the NYPD's Patternizr. *Fordham Urban Law Journal*, Τόμος 47, p. 43.
- [57]: Grote, T. & Berens, P., 2020. On the ethics of algorithmic decision-making in healthcare. *Journal of medical ethics*, 46(3), pp. 205-211.
- [58]: Guidotti, R. και συν., 2018. A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), pp. 1-42.
- [59]: Haeri, M. & Zweig, K., 2020. The crucial role of sensitive attributes in fair classification. *IEEE Symposium Series on Computational Intelligence (SSCI)* , December, pp. 2993-3002.
- [60]: Hajian, S. & Domingo-Ferrer, J., 2012. A methodology for direct and indirect discrimination prevention in data mining. *IEEE transactions on knowledge and data engineering*, 25(7), pp. 1445-1459.
- [61]: Kamiran, F., Žliobaitė, I. & Calders, T., 2013. Quantifying explainable discrimination and removing illegal discrimination in automated decision making. *Knowledge and information systems*, Τόμος 35, pp. 613-644.
- [62]: Kashid, A., Kulkarni, V. & Patankar, R., 2017. Discrimination-aware data mining: a survey. *International Journal of Data Science*, 2(1), pp. 70-84.
- [63]: Kehl, D., Guo, P. & Kessler, S., 2017. *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*, Harvard:

Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School.

- [64]: Kirkpatrick, K., 2017. It's not the algorithm, it's the data. *Communications of the ACM*, 60(2), pp. 21-23.
- [65]: Köchling, A. & Wehner, M., 2020. Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), pp. 795-848.
- [66]: Langer, M. & Landers, R., 2021. The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers. *Computers in Human Behavior*, Τόμος 123, p. 106878.
- [67]: Lawspot, 2021. *Τεχνητή Νοημοσύνη και Ισότητα - Φοιτητές γράφουν για την Τεχνητή Νοημοσύνη.* [Ηλεκτρονικό] Available at: https://www.lawspot.gr/nomika-nea/tehniti-noimosyni-kai-isotita-foitites-grafoyn-gia-tin-tehniti-noimosyni#footnote9_gesuc3d [Πρόσβαση 17 Δεκέμβριος 2022].
- [68]: Le Quy, T. και συν., 2022. A survey on datasets for fairness-aware machine learning. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 12(3), p. e1452.
- [69]: Lee, M., 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), p. 2053951718756684.
- [70]: Lepri, B. και συν., 2018. Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), pp. 611-627.
- [71]: Levy, A., Agrawal, M., Satyanarayan, A. & Sontag, D., 2021. Assessing the impact of automated suggestions on decision making: Domain experts mediate model errors but take less initiative. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

- [72]: Lum, K. & Isaac, W., 2016. To predict and serve?. *Significance*, 13(5), pp. 14-19.
- [73]: Manuel, D. και συν., 2012. Predictive risk algorithms in a population setting: an overview. *Journal of Epidemiology and Community Health*, 66(10), pp. 859-865.
- [74]: Mittelstadt, B. και συν., 2016. The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), p. 2053951716679679.
- [75]: Ntoutsi, E. και συν., 2020. Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), p. e1356.
- [76]: Pedreshi, D., Ruggieri, S. & Turini, F., 2008. Discrimination-aware data mining. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 560-568.
- [77]: Prinsloo, P., 2020. Of ‘black boxes’ and algorithmic decision-making in (higher) education—A commentary. *Big Data & Society*, 7(1), p. 2053951720933994.
- [78]: Selbst, A. & Barocas, S., 2018. The intuitive appeal of explainable machines. *Fordham Law Review*, 87(3), pp. 1085-1139.
- [79]: Starke, C., Baleis, J., Keller, B. & Marcinkowski, F., 2022. Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, 9(2), p. 205395172211151.
- [80]: Sweeney, L., 2013. Discrimination in online ad delivery. *Communications of the ACM*, 56(5), pp. 44-54.
- [81]: Tene, O. & Polonetsky, J., 2017. Taming the Golem: Challenges of ethical algorithmic decision-making. *NCJL & Tech*, Τόμος 19, p. 125.
- [82]: Tolan, S., 2019. Fair and unbiased algorithmic decision making: current state and future challenges. *arXiv preprint arXiv*, Τόμος 1901.04730.
- [83]: Turner, K., 2016. *Meet ‘Ross,’ the newly hired legal robot*. [Ηλεκτρονικό] Available at: <https://www.washingtonpost.com/news/innovations/wp/2016/05/16/meet-ross->

the-newly-hired-legal-robot/

[Πρόσβαση 17 Δεκέμβριος 2022].

- [84]: Vetrò, A., Torchiano, M. & Mecati, M., 2021. A data quality approach to the identification of discrimination risk in automated decision making systems. *Government Information Quarterly*, 38(4), p. 101619.
- [85]: Vetrò, A., Torchiano, M. & Mecati, M., 2021. A data quality approach to the identification of discrimination risk in automated decision making systems. *Government Information Quarterly*, 38(4), p. 101619.
- [86]: Wischmeyer, T., 2020. Artificial intelligence and transparency: opening the black box. Στο: T. Wischmeyer & T. Rademacher, επιμ. *Regulating artificial intelligence*. s.l.:Springer, Cham, pp. 75-101.
- [87]: Xu, R. και συν., 2020. Algorithmic decision making with conditional fairness. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2125-2135.
- [88]: Zhuang, Z. και συν., 2018. Discrimination-aware channel pruning for deep neural networks. *Advances in neural information processing systems*, Τόμος 31.
- [89]: Žliobaitė, I., 2017a. Measuring Discrimination in Algorithmic Decision Making. *Data Mining and Knowledge Discovery*, 31(4), p. 1060–1089.
- [90]: Žliobaitė, I., 2017b. Fairness-aware machine learning: a perspective. *ArXiv*, Τόμος 11636248.
- [91]: Εθνική Επιτροπή για τα Δικαιώματα του Ανθρώπου, 2022. *ΕΕΔΑ, το ανεξάρτητο συμβουλευτικό όργανο της Πολιτείας*. [Ηλεκτρονικό] Available at: <https://www.nchr.gr/eeda.html:%20%CE%95%CE%95%CE%94%CE%91,%20%CF%84%CE%BF%20%CE%B1%CE%BD%CE%B5%CE%BE%CE%A%CF%81%CF%84%CE%B7%CF%84%CE%BF%20%CF%83%CF%85%CE%BC%CE%B2%CE%BF%CF%85%CE%BB%CE%B5%CF%85%CF%84%CE%B9%CE%BA%CF%8C%20%CF%8C%CF%81%CE%B3%CE%B1>

%CE%BD%CE%B

[Πρόσβαση 12 Δεκέμβριος 2022].