



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS



Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας
Πρόγραμμα Μεταπτυχιακών Σπουδών
«ΨΗΦΙΑΚΟΣ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΦΟΡΟΛΟΓΙΚΩΝ
ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΩΝ ΥΠΗΡΕΣΙΩΝ»

Διπλωματική Εργασία

«Εφαρμογή Νέων Τεχνολογιών Εξαγωγής Πληροφορίας από Διαδικτυακές Αγγελίες
Εργασίας»

Θωμαΐ Χριστίνα

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής Γιωτόπουλος Κωνσταντίνος	
Α΄ Συν-Επιβλέπων Χαλκιάπουλος Κωνσταντίνος	Β΄ Συν-Επιβλέπων Θεοδωρακόπουλος Λεωνίδα

Πάτρα, Ιούνιος 2024

© Copyright συγγραφής ...Θωμαΐ Χριστίνα 2024

Με την επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας δεν συνεπάγεται απαραίτητως και αποδοχή των απόψεων του συγγραφέως εκ μέρους του τμήματος.

Σελίδα Αφιέρωσης

Με την παρούσα εργασία, ολοκληρώνεται ο κύκλος των σπουδών μου στο Μεταπτυχιακό Πρόγραμμα «Ψηφιακός Μετασχηματισμός Φορολογικών και Χρηματοοικονομικών Υπηρεσιών» του τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του Πανεπιστημίου Πατρών. Πριν την ανάλυση του θέματος της παρούσας διπλωματικής εργασίας, νιώθω την υποχρέωση να ευχαριστήσω τους ανθρώπους που έπαιξαν σημαντικό ρόλο κατά τη διάρκεια εκπόνησης της εργασίας μου.

Πρώτα από όλα, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας τον κ. Γιωτόπουλο Κωνσταντίνο για την πολύτιμη βοήθεια που μου πρόσφερε για την εκπόνηση της διπλωματικής. Καθώς, θα ήθελα να τον ευχαριστώ για την εμπιστοσύνη που μου έδειξε για την ανάθεση της διπλωματικής εργασίας ,για όλα όσα μας δίδαξε, για το επιστημονικό υλικό που μου πρόσφερε, τις συμβουλές του, την υπομονή του και τις ώρες που αφιέρωσε για την καθοδήγηση μου ώστε να καταφέρω να ολοκληρώσω με επιτυχία την διπλωματική μου. Στο σημείο αυτό θα ήθελα ευχαριστήσω θερμά τον κύριο Νίκο Ζώτο για την παροχή συμβουλευτικής υποστήριξης σε όλα τα στάδια εκπόνησης της εργασίας με συγκεκριμένες παρατηρήσεις, στόχους και για την πολύτιμη καθοδήγηση του.

Ιδιαίτερα θα ήθελα να ευχαριστήσω την οικογένεια μου για την οικονομική και ψυχολογική υποστήριξη. Καθώς , και για την συνεχείς συμπαράσταση, για τις συμβουλές τους και για όλα όσα μου έχουν προσφέρει σε όλα τα χρόνια της ζωής για την ολοκλήρωση των ακαδημαϊκών σπουδών μου.

Τέλος, θα ήθελα να ευχαριστήσω θερμά τους φίλους μου για την ανοχή και την κατανόηση που επέδειξαν όλο το διάστημα της εκπόνησης της διπλωματικής εργασίας.

Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει τη σημασία της ανάλυσης της αγοράς εργασίας μέσω διαδικτυακών αγγελιών, επικεντρώνοντας στην εξαγωγή και επεξεργασία πληροφοριών με τη χρήση προηγμένων τεχνολογιών. Ο σκοπός της έρευνας είναι η ανάπτυξη και εφαρμογή προηγμένων συστημάτων και μεθόδων για την αυτόματη ανάλυση και κατηγοριοποίηση δεδομένων από διαδικτυακές αγγελίες εργασίας. Η εργασία αξιοποιεί τις δυνατότητες της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης για τη δημιουργία αυτόνομων συστημάτων που μπορούν να αντιμετωπίσουν τα προβλήματα που σχετίζονται με την ανάλυση και επεξεργασία αυτών των δεδομένων.

Η μεθοδολογία περιλαμβάνει έναν συνδυασμό θεωρητικής κατανόησης και εφαρμοσμένης μεθοδολογίας. Αρχικά, το πρόβλημα προσδιορίστηκε με σαφήνεια, επικεντρώνοντας στην ανάγκη ανάλυσης των δυναμικών της αγοράς εργασίας μέσω διαδικτυακών αγγελιών. Στη συνέχεια, εξετάστηκαν οι σύγχρονες τεχνολογικές εξελίξεις στην εξαγωγή πληροφοριών, συμπεριλαμβανομένων των μοντέλων *ML* και *AI*, καθώς και των αλγορίθμων Επεξεργασίας Φυσικής Γλώσσας (*NLP*). Η εφαρμοσμένη μεθοδολογία περιλαμβάνει τη χρήση εργαλείων όπως *APIs*, *Web Scraping* και ανάλυση με τη χρήση μοντέλων *ML* και *AI*. Η ανάλυση δείχνει ότι η χρήση τεχνολογιών *ML* και *AI* μπορεί να βελτιώσει σημαντικά την εξαγωγή και επεξεργασία πληροφοριών από διαδικτυακές αγγελίες εργασίας. Τα συστήματα που αναπτύχθηκαν είναι σε θέση να κατηγοριοποιούν και να αναλύουν τις αγγελίες με υψηλή ακρίβεια, παρέχοντας πολύτιμες πληροφορίες για τις απαιτούμενες δεξιότητες και τις διαρθρωτικές ανάγκες της αγοράς εργασίας.

Ένας από τους κύριους περιορισμούς της έρευνας είναι η φύση των μοντέλων *ML* ως "μαύρα κουτιά", που καθιστούν δύσκολη την κατανόηση του τρόπου λήψης αποφάσεων. Επιπλέον, η ανεπάρκεια τυποποίησης και η ύπαρξη θορύβων στα δεδομένα από τις αγγελίες αποτελούν προκλήσεις για την ακριβή ανάλυση. Οι πρακτικές εφαρμογές της έρευνας περιλαμβάνουν τη χρήση των αναπτυγμένων συστημάτων για την υποστήριξη επιχειρήσεων και οργανισμών στην κατανόηση των αναγκών της αγοράς εργασίας και στη λήψη αποφάσεων σχετικά με την πρόσληψη προσωπικού.

Τέλος, η έρευνα είναι πρωτότυπη στην προσέγγισή της, συνδυάζοντας τεχνολογίες *ML*, *AI* και *NLP* για την ανάλυση διαδικτυακών αγγελιών εργασίας. Επίσης, η εργασία αναδεικνύει

τις μελλοντικές προοπτικές των Μεγάλων Γλωσσικών Μοντέλων (*LLMs*) στην εξαγωγή πληροφοριών, καθώς και τη σημασία της εξηγήσιμης ΑΙ (*Explainable AI*) για τη βελτίωση της διαφάνειας και της κατανόησης των αποτελεσμάτων των μοντέλων.

Λέξεις Κλειδιά: Εφαρμογή νέων τεχνολογιών, Εξαγωγή πληροφορίας, Διαδικτυακές αγγελίες εργασίας

Abstract

This thesis examines the importance of analysing the labour market through online advertisements, focusing on the extraction and processing of information using advanced technologies. The aim of the research is to develop and apply advanced systems and methods for automatic analysis and categorization of data from online job advertisements. The work exploits the capabilities of Machine Learning and Artificial Intelligence to create autonomous systems that can address the problems associated with the analysis and processing of this data.

The methodology involves a combination of theoretical understanding and applied methodology. Initially, the problem was clearly defined, focusing on the need to analyze labor market dynamics through online job postings. Then, current technological developments in information extraction, including ML and AI models, as well as Natural Language Processing (NLP) algorithms, were examined. The applied methodology includes the use of tools such as APIs, Web Scraping and analysis using ML and AI models. The analysis shows that the use of ML and AI technologies can significantly improve the extraction and processing of information from online job advertisements. The developed systems are able to categorize and analyze the advertisements with high accuracy, providing valuable information about the required skills and structural needs of the labor market.

One of the main limitations of the research is the nature of ML models as 'black boxes', making it difficult to understand how decisions are made. In addition, the lack of standardisation and the presence of noise in the advertisement data pose challenges to accurate analysis. Practical applications of the research include the use of the developed systems to support firms and organisations in understanding labour market needs and making recruitment decisions.

Finally, the research is original in its approach, combining ML, AI and NLP technologies to analyse online job advertisements. Also, the paper highlights the future potential of Large Language Models (LLMs) in information extraction, as well as the importance of Explainable AI to improve the transparency and understanding of model results.

Keywords: Application of new technologies, Information extraction, Online job posting.

Πίνακας Περιεχομένων

Σελίδα Αφιέρωσης	iii
Περίληψη	i
Abstract	iii
Πίνακας Συντομογραφιών	vi
Πρόλογος.....	vii
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	1
1.1.Προσδιορισμός προβλήματος ,σκοπού και στόχου διπλωματικής.....	1
1.2.Η προηγούμενη γνώση στην προσέγγιση του προβλήματος	1
1.3.Περιγραφή μεθοδολογίας	2
1.4. Δομή της εργασίας	3
ΚΕΦΑΛΑΙΟ 2 : ΤΕΧΝΟΛΟΓΙΕΣ ΕΞΑΓΩΓΗΣ ΠΛΗΡΟΦΟΡΙΩΝ	3
2.1. Η εξέλιξη της εξαγωγής πληροφοριών	3
2.1.1. Η σημασία της έννοιας εξαγωγής πληροφορίας	5
2.1.2. Εργαλεία εξαγωγής πληροφορίας	7
2.1.3. Τα πλεονεκτήματα της εξαγωγής πληροφορίας	9
2.2. Τεχνικές εξαγωγής πληροφορίας από διαδικτυακές αγγελίες εργασίας	11
2.2.1. Η χρήση των APIs των ιστότοπων αγγελιών εργασίας	12
2.2.2. Web Scrapping.....	15
2.2.3. Η χρήση των RSS Feeds	18
2.3. Προβλήματα που αντιμετωπίζονται στην ανάλυση και επεξεργασία πληροφοριών απο διαδικτυακές αγγελιών εργασίας	20
ΚΕΦΑΛΑΙΟ 3: Η ΑΝΑΠΤΥΞΗ ΣΥΣΤΗΜΑΤΩΝ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΑΓΓΕΛΙΩΝ ΕΡΓΑΣΙΑΣ	23
3.1. Χρήση μοντέλων Machine Learning και της τεχνητής νοημοσύνης για την κατηγοριοποίηση και ανάλυση των διαδικτυακών αγγελιών εργασίας	23
3.1.1. Ορισμός της Μηχανικής Μάθησης.....	24
3.1.2. Η λειτουργία της τεχνολογίας της Μηχανικής Μάθησης	26
3.1.3. Τα προνόμια που προσφέρει η Μηχανική Μάθηση στις μέρες μας	30
3.1.4. Η χρήση της μηχανικής μάθησης στην εξαγωγή πληροφορίες από διαδικτυακές αγγελίες εργασίας.....	33
3.2.Ανάλυση αγγελιών εργασίας με την χρήση της τεχνητής νοημοσύνης.....	33
3.2.1. Η σημασία τις τεχνητής νοημοσύνης	37
3.2.2. Πλεονεκτήματα και Μειονεκτήματα της Τεχνητής Νοημοσύνης	39

3.2.3. Η τεχνητή νοημοσύνης διευκολύνει την εξαγωγή πληροφορίας από διαδικτυακές αγγελίες εργασίας.....	44
ΚΕΦΑΛΑΙΟ 4: Η ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΚΑΙ ΤΕΧΝΙΚΩΝ ΕΞΑΓΩΓΗΣ ΠΛΗΡΟΦΟΡΙΑΣ	46
4.1. Η έννοια της φυσικής γλώσσας επεξεργασίας (NPL) και η εφαρμογή στην αφαίρεση πληροφορίας	46
4.1.1. Εξαγωγή πληροφορίας με την χρήση των τεχνικών NPL.....	47
4.2. Η ανάλυση μεθοδολογιών Machine Learning στην εξαγωγή πληροφορίας	48
4.3. Μελλοντικές προοπτικές για τη χρήση των LLMs στην εξαγωγή πληροφοριών.....	50
Συμπεράσματα.....	52
ΒΙΒΛΙΟΓΡΑΦΙΑ	54

Πίνακας Συντομογραφιών

AI	Artificial Intelligence
API	Application Programming Interface
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
IE	Information Extraction
IP	Internet Protocol
LLMs	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
NLTK	Natural Language Toolkit
RDF	Resource Description Framework
SEO	Search Engine Optimization
SVM	Support Vector Machines
TN	Τεχνητή Νοημοσύνη
URL	Uniform Resource Locator
XML	Extensible Markup Language

Πρόλογος

Η παρούσα διπλωματική εργασία αναδεικνύει τη σημασία της ανάλυσης της αγοράς εργασίας μέσω των διαδικτυακών αγγελιών, επικεντρώνοντας στην εξαγωγή και επεξεργασία πληροφοριών με τη χρήση προηγμένων τεχνολογιών. Η ανάλυση αυτή αποτελεί κρίσιμο εργαλείο για την κατανόηση των διαρθρωτικών χαρακτηριστικών της αγοράς εργασίας, των επιχειρηματικών αναγκών, και των δεξιοτήτων που απαιτούνται σε διάφορους κλάδους.

Η εργασία διακρίνεται σε τέσσερα κεφάλαια, κάθε ένα εξερευνά διαφορετικές πτυχές της ανάλυσης των διαδικτυακών αγγελιών εργασίας και της εφαρμογής τεχνολογιών εξαγωγής και επεξεργασίας πληροφοριών. Στο πρώτο κεφάλαιο παρατίθεται ο σκοπός, ο στόχος, και ο προβληματισμός που οδήγησε στην εκπόνηση της διπλωματικής. Στην συνέχεια, στο δεύτερο κεφάλαιο εστιάζει στις τεχνολογίες εξαγωγής πληροφοριών, αναδεικνύοντας τη σημασία της εξαγωγής πληροφοριών από διαδικτυακές αγγελίες εργασίας, και εξετάζοντας τις τεχνικές όπως τη χρήση *APIs*, το *Web Scraping* και τα *RSS Feeds*. Έπειτα, το τρίτο κεφάλαιο εστιάζει στην ανάπτυξη συστημάτων για την ανάλυση των αγγελιών εργασίας, εξερευνώντας τη χρήση μοντέλων *Machine Learning* και τεχνητής νοημοσύνης. Τέλος, στο τέταρτο κεφάλαιο αναλύουμε και τεχνικές εξαγωγής πληροφορίας όπως οι τεχνικές *NPL*, *Machine Learning* και αναφέρουμε μελλοντικές προοπτικές *LLMs*.

Μέσα από αυτά τα κεφάλαια, προσεγγίζουμε την πολυπλοκότητα της ανάλυσης των διαδικτυακών αγγελιών εργασίας, αποκαλύπτοντας τις τεχνολογικές και αναλυτικές διαδικασίες που απαιτούνται για την αποτελεσματική εξαγωγή και επεξεργασία πληροφοριών. Η μεθοδολογία που ακολουθήθηκε βασίστηκε σε έρευνα και ανάλυση πληθώρας επιστημονικών πηγών.

Τέλος, θα ήθελα να εκφράσω τις ευχαριστίες μου σε προς όλους εκείνους που συνέβαλαν στην υλοποίηση αυτής της εργασίας, συμπεριλαμβανομένων του επιβλέποντος καθηγητή, της εξεταστικής επιτροπής, των δημόσιων υπηρεσιών, οργανισμών, εταιριών, καθηγητών, συμφοιτητών και άλλων ατόμων που παρείχαν στήριξη και συνεργασία στη διάρκεια αυτής της εργασίας.

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1.Προσδιορισμός προβλήματος ,σκοπού και στόχου διπλωματικής

Ο προσδιορισμός του προβλήματος που εξετάζει η διπλωματική εργασία επικεντρώνεται στην ανάλυση του σύγχρονου τοπίου της απασχόλησης μέσω των διαδικτυακών αγγελιών εργασίας. Παρουσιάζει την ανάγκη για την ανάπτυξη και την εφαρμογή προηγμένων τεχνολογιών και μεθόδων, προκειμένου να αξιοποιηθεί αποτελεσματικά η πληθώρα πληροφοριών που παρέχονται από τις πλατφόρμες αυτές. Καθιστά, σαφές τα προβλήματα που προκύπτουν κατά την ανάλυση και την επεξεργασία των δεδομένων από τις διαδικτυακές αγγελίες εργασίας, όπως η ανεπάρκεια τυποποίησης και η ύπαρξη ανεπιθύμητων θορύβων στα δεδομένα.

Ο σκοπός και οι στόχοι της διπλωματικής εργασίας εστιάζουν στην ανάπτυξη και την εφαρμογή προηγμένων συστημάτων και μεθόδων, με στόχο την αυτόματη ανάλυση και κατηγοριοποίηση των δεδομένων από διαδικτυακές αγγελίες εργασίας. Αξιοποιεί τις δυνατότητες της Μηχανικής Μάθησης και της Τεχνητής Νοημοσύνης για την ανάπτυξη αυτόνομων συστημάτων που θα είναι σε θέση να αντιμετωπίζουν τα προβλήματα που παρουσιάζονται στην ανάλυση και επεξεργασία των δεδομένων αυτών. Σκοπός της είναι επίσης, η ανάπτυξη Αλγορίθμων Επεξεργασίας Φυσικής Γλώσσας για την ανάλυση και την εξαγωγή σημαντικών πληροφοριών.

1.2.Η προηγούμενη γνώση στην προσέγγιση του προβλήματος

Η προηγούμενη γνώση στην προσέγγιση του προβλήματος της ανάλυσης και επεξεργασίας πληροφοριών από διαδικτυακές αγγελίες εργασίας είναι πολύτιμη για την κατανόηση των προκλήσεων και των πιθανών λύσεων που προσφέρονται. Η εξέλιξη της τεχνολογίας στον τομέα της εξαγωγής πληροφοριών από διαδικτυακούς πόρους, όπως οι ιστοσελίδες αγγελιών εργασίας, έχει διευκολύνει τη διαδικασία ανάλυσης και επεξεργασίας μεγάλου όγκου δεδομένων. Η χρήση τεχνικών όπως η εξαγωγή δεδομένων μέσω *APIs*, το *Web Scraping* και η χρήση *RSS Feeds* έχουν ανοίξει νέες προοπτικές για την ανάλυση των αγγελιών

εργασίας. Παράλληλα, η εφαρμογή της τεχνητής νοημοσύνης και της μηχανικής μάθησης έχει επιτρέψει την αυτοματοποίηση και την ανάλυση των δεδομένων, ενισχύοντας την αποτελεσματικότητα και την ακρίβεια των αναλύσεων. Η εξέλιξη της φυσικής γλώσσας επεξεργασίας και η εφαρμογή αλγορίθμων *NLP* στην ανάλυση κειμένου από τις αγγελίες εργασίας αποτελούν επίσης σημαντικά εργαλεία για την ανάκτηση χρήσιμων πληροφοριών και την αυτοματοποίηση της επεξεργασίας τους. Με βάση αυτήν την προηγούμενη γνώση, οι ερευνητές μπορούν να προσεγγίσουν το πρόβλημα αυτό με την αξιοποίησή των κατάλληλων τεχνικών.

1.3. Περιγραφή μεθοδολογίας

Η μεθοδολογία που ακολουθήθηκε στην ανάπτυξη του θέματος της εργασίας ενσωμάτωσε τον συνδυασμό προβληματισμού, θεωρητικής κατανόησης και εφαρμοσμένης μεθοδολογίας. Αρχικά, το πρόβλημα προσδιορίστηκε με σαφήνεια, εστιάζοντας στην ανάγκη να αναλυθούν οι δυναμικές της αγοράς εργασίας μέσω των διαδικτυακών αγγελιών. Στη συνέχεια, η θεωρητική προσέγγιση εξέτασε τις σύγχρονες τεχνολογικές εξελίξεις στον τομέα της εξαγωγής πληροφοριών, συμπεριλαμβανομένων των μοντέλων *Machine Learning* και της τεχνητής νοημοσύνης καθώς και των αλγορίθμων *NLP*. Αυτή η θεωρητική βάση κατέστησε δυνατή την κατανόηση των συστημάτων και των μεθόδων που χρησιμοποιήθηκαν για την ανάπτυξη της έρευνας.

Η εφαρμοσμένη μεθοδολογία περιλαμβάνει τη χρήση συγκεκριμένων εργαλείων και τεχνικών, όπως η χρήση *APIs*, το *Web Scraping*, η ανάλυση με χρήση μοντέλων *Machine Learning* και τεχνητής νοημοσύνης. Η εφαρμογή αυτών των μεθόδων αναδείχθηκε ως κρίσιμη για την επεξεργασία και ανάλυση των δεδομένων από διαδικτυακές αγγελίες εργασίας. Τέλος, η εφαρμογή αλγορίθμων *NLP* επέτρεψε στην ανάλυση του περιεχομένου των αγγελιών, να εξάγουν χρήσιμες πληροφορίες για την κατηγοριοποίηση και την ανάλυση των δεδομένων. Αυτή η συνδυασμένη προσέγγιση παρείχε μια ολοκληρωμένη μέθοδο ανάπτυξης και αξιολόγησης του θέματος, που συνέβαλε στην πλήρη κατανόηση και αξιολόγηση των δεδομένων της έρευνας.

1.4. Δομή της εργασίας

Η παρούσα εργασία διαρθρώνεται σε τέσσερα κεφάλαια. Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή σχετικά με τον προσδιορισμό του προβλήματος, τους στόχους και τους σκοπούς της διπλωματικής εργασίας. Ύστερα, αναφέρεται στην προηγούμενη γνώση στην προσέγγιση του προβλήματος που θέσαμε. Τέλος, παραθέσαμε την περιγραφή μεθοδολογίας / μεθόδου που χρησιμοποιήθηκε για την ανάπτυξη της παρούσας διπλωματικής .

Το δεύτερο κεφάλαιο επικεντρώνεται στις τεχνολογίες εξαγωγής πληροφοριών από διαδικτυακές αγγελίες εργασίας. Αρχικά, αναλύει την εξέλιξη της έννοιας της εξαγωγής πληροφοριών, εστιάζοντας στη σημασία της και στα εργαλεία που χρησιμοποιούνται για την εκτέλεσή της. Στη συνέχεια, παρουσιάζει τις τεχνικές εξαγωγής πληροφοριών από διαδικτυακές αγγελίες εργασίας, συμπεριλαμβανομένης της χρήσης των *APIs* των ιστότοπων αγγελιών εργασίας, του *Web Scraping* και των *RSS Feeds*. Τέλος, αναδεικνύει τα προβλήματα που αντιμετωπίζονται στην ανάλυση και επεξεργασία πληροφοριών από διαδικτυακές αγγελίες εργασίας .

Στο τρίτο κεφάλαιο, εξετάζεται η ανάπτυξη συστημάτων για την αναζήτηση και ανάλυση αγγελιών εργασίας με χρήση μοντέλων *Machine Learning*. Πρώτα, εξηγείται η έννοια της Μηχανικής Μάθησης και η λειτουργία της, ενώ στη συνέχεια εξετάζονται τα οφέλη που μας προσφέρει στην σημερινή εποχή και η εφαρμογή της στην εξαγωγή πληροφοριών από διαδικτυακές αγγελίες εργασίας. Στη συνέχεια, παρουσιάζεται η ανάλυση αγγελιών εργασίας με τη χρήση της τεχνητής νοημοσύνης, εξετάζοντας τη σημασία της, τα πλεονεκτήματα και μειονεκτήματά της, καθώς και τον τρόπο με τον οποίο διευκολύνει την εξαγωγή πληροφοριών από διαδικτυακές αγγελίες εργασίας.

Το τέταρτο κεφάλαιο εστιάζει στην εφαρμογή αλγορίθμων και τεχνικές εξαγωγής πληροφορίας .Πιο συγκεκριμένα, αναλύουμε την έννοια και τεχνικές *NPL*, αναφέρουμε τις τεχνικές *Machine Learning* και τέλος, αναπτύσσουμε τις μελλοντικές τεχνικές *LLMs*.

ΚΕΦΑΛΑΙΟ 2 : ΤΕΧΝΟΛΟΓΙΕΣ ΕΞΑΓΩΓΗΣ ΠΛΗΡΟΦΟΡΙΩΝ

2.1. Η εξέλιξη της εξαγωγής πληροφοριών

Η εξαγωγή πληροφοριών (*Information Extraction*) έχει γίνει ένα απαραίτητο εργαλείο καθώς, παρέχει την ικανότητά συλλογής δομημένης γνώσης από τον συνεχώς αυξανόμενο

πλήθος αδόμητων δεδομένων. Αυτοματοποιεί τη διαδικασία εξαγωγής συγκεκριμένων, δομημένων πληροφοριών από διάφορες πηγές κειμένου. Ο τομέας αυτός έχει γνωρίσει σημαντικές εξελίξεις, μετασχηματίζοντας τον τρόπο αλληλεπίδρασης των πληροφοριών και τον τρόπο που τις αξιοποιούμε. Αυτή η ανασκόπηση εμβαθύνει στην ιστορική εξέλιξη του ΙΕ, διερευνώντας τις ρίζες του και τις σημαντικές εξελίξεις που έχουν διαμορφώσει αυτό το κρίσιμο πεδίο.

Την δεκαετία του 1950 και του 1960 μόλις είχαν αρχίσει οι έρευνες για την τεχνητή νοημοσύνη. Όμως, η πρωτοποριακή εργασία στην επεξεργασία φυσικής γλώσσας έθεσε τις βάσεις για την κατανόηση και τον χειρισμό δεδομένων κειμένου. Συστήματα όπως το *LUNAR*, που αναπτύχθηκαν τη δεκαετία του 1960 για γεωλογική εξερεύνηση, χρησιμοποίησαν κωδικοποιημένους κανόνες για τον εντοπισμό συγκεκριμένων οντοτήτων και σχέσεων εντός επιστημονικών εγγράφων. Αυτά τα πρώιμα συστήματα βασίζονταν σε μεγάλο βαθμό σε ειδικές γνώσεις για τον τομέα και ήταν περιορισμένες στην ικανότητά τους να χειρίζονται παραλλαγές στη γλώσσα.

Στη δεκαετία του 1990 σημειώθηκε μια στροφή προς τις στατιστικές προσεγγίσεις. Τεχνικές όπως *n-grams* (ακολουθίες λέξεων) και κρυφά μοντέλα *Markov*. Υιοθετήθηκαν για την αναγνώριση ονομαστικών οντοτήτων (*NER*), μια βασική εργασία του ΙΕ που περιλαμβάνει τον εντοπισμό και την ταξινόμηση ονομασμένων οντοτήτων όπως άτομα, οργανισμοί και τοποθεσίες. Η σειρά *Message Understanding Conference (MUC)*, που ξεκίνησε το 1991, έπαιξε σημαντικό ρόλο στην αξιολόγηση και την προώθηση των συστημάτων *NER*.

Ο 21ος αιώνας αποτελεί μια νέα εποχή με την άνοδο των τεχνικών μηχανικής μάθησης. Οι εποπτευόμενοι αλγόριθμοι μάθησης που εκπαιδεύτηκαν σε μεγάλα σχολιασμένα σύνολα δεδομένων πέτυχαν σημαντικές βελτιώσεις στο *NER*, στην εξαγωγή σχέσεων και στον προσδιορισμό των σχέσεων μεταξύ οντοτήτων. Οι διανυσματικές μηχανές υποστήριξης (*SVM*) και τα τυχαία πεδία υπό όρους (*CRF*) έγιναν δημοφιλή μοντέλα για αυτές τις εργασίες.

Την τελευταία δεκαετία έχει κυριαρχήσει η χρήση της βαθιάς μάθησης, ιδιαίτερα των συνελκτικών νευρωνικών δικτύων (*CNN*) και των επαναλαμβανόμενων νευρωνικών δικτύων (*RNN*) με μονάδες μακροπρόθεσμης βραχυπρόθεσμης μνήμης (*LSTM*). Αυτά τα ισχυρά μοντέλα μπορούν να αποτυπώσουν πολύπλοκα μοτίβα από τεράστιες ποσότητες δεδομένων κειμένου, οδηγώντας σε σημαντικές προόδους στην ακρίβεια και την ευρωστία. Τα αμφίδρομα *LSTM*, για παράδειγμα, διαπρέπουν στον χειρισμό εξαρτήσεων μεγάλων αποστάσεων στο κείμενο, κάτι που είναι ζωτικής σημασίας για την κατανόηση των σχέσεων μεταξύ οντοτήτων.

Παρά την αξιοσημείωτη πρόοδο, η ΙΕ εξακολουθεί να αντιμετωπίζει προκλήσεις. Ο χειρισμός της ασάφειας, η κατανόηση πολύπλοκων δομών προτάσεων και η αντιμετώπιση

περιορισμένων ή θορυβωδών δεδομένων παραμένουν εμπόδια. Η εξαγωγή πληροφορίας εφαρμόζεται σε ένα ευρύ φάσμα τύπων κειμένου καθώς, παρουσιάζει σημαντική πρόκληση σε σύγκριση με συστήματα που αφορούν συγκεκριμένους τομείς. Ο τομέας της εξαγωγής πληροφοριών συνεχίζει να εξελίσσεται ραγδαία. Με τις εξελίξεις στη βαθιά εκμάθηση και τη διαρκώς αυξανόμενη διαθεσιμότητα δεδομένων κειμένου, μπορούμε να περιμένουμε ακόμη πιο εξελιγμένα συστήματα *IE* ικανά να χειρίζονται πολύπλοκες εργασίες εξαγωγής πληροφοριών σε διάφορους τομείς. Αυτό αναμφίβολα θα διαδραματίσει ζωτικό ρόλο στο ξεκλείδωμα των πραγματικών δυνατοτήτων του τεράστιου όγκου πληροφοριών κειμένου που παράγεται στην ψηφιακή εποχή μας. (Jurafsky, 2021)

2.1.1. Η σημασία της έννοιας εξαγωγής πληροφορίας

Οι πληροφορίες που προσφέρουν τα δεδομένα αποτελούν πολύτιμο περιεχόμενο για την ανάλυση και την κατανόηση διαφόρων φαινομένων, συμπεριλαμβανομένων των κοινωνικών, οικονομικών και επιστημονικών πτυχών. Μέσω των δεδομένων, μπορούμε να ανιχνεύσουμε πρότυπα, τάσεις και ανωμαλίες που μπορούν να μας βοηθήσουν στη λήψη αποφάσεων και στην εξαγωγή συμπερασμάτων. Η αξιοποίηση των δεδομένων απαιτεί όμως προσεκτική επεξεργασία και καθαρισμό, καθώς η ακρίβεια και η αξιοπιστία των αναλύσεων εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων που χρησιμοποιούμε. Η επεξεργασία και ο καθαρισμός των δεδομένων αποτελούν κρίσιμα βήματα πριν από την ανάλυση και την ερμηνεία τους. Κατά τη διαδικασία αυτή, εκτελούνται διάφορες δράσεις όπως η ανίχνευση, η αντιμετώπιση απουσιαζουσών τιμών, η αφαίρεση διπλοεγγραφών, η αντιμετώπιση ανωμαλιών, εκτροπών, και η εναρμόνιση των δεδομένων από διαφορετικές πηγές. Ο στόχος είναι να επιτευχθεί η μείωση του θορύβου, η βελτίωση της αξιοπιστίας των δεδομένων, προετοιμάζοντάς για περαιτέρω ανάλυση και εφαρμογή μοντέλων. Η διαδικασία αυτή απαιτεί τόσο τεχνικές γνώσεις όσο και καλή κατανόηση του πεδίου των δεδομένων που εξετάζουμε.

Την σήμερον ημέρα τα δεδομένα αναδεικνύονται ως η πιο πολύτιμη πρώτη ύλη για την κατανόηση και τη λήψη αποφάσεων. Η αξία των δεδομένων δεν είναι απολύτως προφανής εάν δεν υποστούν επεξεργασία και ανάλυση. Η ανάλυση δεδομένων αποτελεί έναν σημαντικό τομέα της σύγχρονης επιστήμης και τεχνολογίας που συνδυάζει την τεχνολογία των δεδομένων με την επιστήμη της πληροφορίας. Τα δεδομένα, ως πηγή πληροφοριών, με την κατάλληλη

επεξεργασία και ανάλυση, μπορούν να αποκαλύψουν πρότυπα, τάσεις και συσχετίσεις. Μέσω της εξαγωγής πληροφοριών από τα δεδομένα, είναι δυνατόν να αναδειχθούν οι πλέον σημαντικές πληροφορίες και να ληφθούν αποφάσεις με βάση αυτές. Επομένως, η αποτελεσματική εξαγωγή πληροφοριών από τα δεδομένα αποτελεί κρίσιμο βήμα για την αξιοποίησή τους σε διάφορους τομείς, όπως η επιστημονική έρευνα, η επιχειρηματική λήψη αποφάσεων και η κοινωνική ανάπτυξη.

Ο ρόλος της Εξαγωγής πληροφοριών, στα πλαίσια της ανάκτησης πληροφοριών και διαχείρισης γνώσης, είναι η αναγνώριση εξειδικευμένης πληροφορίας και η εξαγωγή γνώσης από μη δομημένα δεδομένα με μηχανικό (αυτόματο) τρόπο. Αντίθετα, με την κλασική ανάκτηση πληροφοριών, σύμφωνα με την οποία η αναζήτηση γίνεται με βάση συγκεκριμένες λέξεις-κλειδιά και το αποτέλεσμα περιλαμβάνει μόνο κείμενα στα οποία βρίσκεται (ενδεχομένως) η χρήσιμη πληροφορία, η εξόρυξη πληροφοριών στοχεύει ακριβώς στην αναγνώριση της χρήσιμης μόνο πληροφορίας. Δεδομένου του μεγάλου όγκου πληροφοριών που παράγονται και διακινούνται σήμερα στο διαδίκτυο, το ζητούμενο στις μέρες μας είναι όχι απλώς η κατοχή της πληροφορίας αλλά η διαχείριση της πληροφορίας και ο εντοπισμός της «σχετικής» πληροφορίας.

Η εξαγωγή ή εξόρυξη πληροφοριών είναι ένα είδος ανάκτησης πληροφοριών από μη δομημένα ή ημιδομημένα δεδομένα (*unstructured or semi-structured data* – τυπικά κείμενα γραμμένα σε φυσική γλώσσα, ιστοσελίδες κ.α.) και αποτελεί κλάδο της σύγχρονης επιστήμης των υπολογιστών. Αποτελεί τη διαδικασία συλλογής ή ανάκτησης διαφορετικών τύπων δεδομένων από διάφορες πηγές, πολλές από τις οποίες μπορεί να είναι κακώς οργανωμένες ή εντελώς αδόμητες. Η εξαγωγή δεδομένων καθιστά δυνατή την ενοποίηση, την επεξεργασία και τη βελτίωση των δεδομένων, ώστε να μπορούν να αποθηκευτούν σε μια κεντρική τοποθεσία προκειμένου να μετασχηματιστούν. (Βικιπαίδεια, 2023).

Η εποχή του διαδικτύου συνεχίζει να αναπτύσσεται με αποτέλεσμα να παράγεται ένας ασύλληπτος όγκος δεδομένων. Εκτιμάται ότι το 2025, τα δεδομένα που θα κυκλοφορούν στο διαδίκτυο θα φτάσουν 163 *zettabytes*. Έτσι, λόγω των τεράστιων ποσοτήτων και της πολυπλοκότητας των μη δομημένων δεδομένων, θα ήταν σχεδόν αδύνατο να εξάγουμε χειροκίνητα τις σχετικές πληροφορίες από όλα τα δεδομένα. Η ύπαρξη ενός αυτοματοποιημένου τρόπου εξαγωγής πληροφοριών από διάφορες μορφές δεδομένων, ιδίως μη δομημένων δεδομένων, και η μετέπειτα παρουσίαση των πληροφοριών αυτών με δομημένο τρόπο επιφέρει πολλά οφέλη. Όπως, είναι η σημαντική μείωση του χρόνου όπου δαπανάται για την εξαγωγή των πληροφοριών. Η εξέλιξη της τεχνολογίας στα συστήματα εξαγωγής

πληροφοριών μπορούν να εκτελέσουν αυτό το έργο με σημαντικά ταχύτερο ρυθμό από ό,τι οι άνθρωποι. (Egnati Simply Intelligent , 2021)

2.1.2. Εργαλεία εξαγωγής πληροφορίας

Τα εργαλεία εξαγωγής πληροφοριών είναι εφαρμογές λογισμικού που έχουν σχεδιαστεί για την αποτελεσματική συλλογή συγκεκριμένων σημείων δεδομένων από διάφορες πηγές. Τα εργαλεία αυτά διαδραματίζουν κρίσιμο ρόλο στο σύγχρονο τοπίο που βασίζεται στα δεδομένα, επιτρέποντας στις επιχειρήσεις να εξορθολογήσουν τις διαδικασίες, να αποκτήσουν πολύτιμες πληροφορίες και να λαμβάνουν τεκμηριωμένες αποφάσεις. Οι κύριες κατηγορίες δεδομένων είναι οι εξής:

- a. **Δομημένα δεδομένα:** Πρόκειται για πληροφορίες οργανωμένες σε προκαθορισμένη μορφή, όπως πίνακες σε λογιστικά φύλλα ή βάσεις δεδομένων. Η τεχνολογία οπτικής αναγνώρισης χαρακτήρων (*OCR*) χρησιμοποιείται συνήθως σε αυτό το πλαίσιο για τη μετατροπή σαρωμένων εγγράφων σε κείμενο αναγνώσιμο από μηχανήματα. (Astera Software, 2023)
- b. **Ημιδομημένα δεδομένα:** Περιλαμβάνουν πληροφορίες με μερικώς καθορισμένη δομή, όπως ιστοσελίδες με ετικέτες *HTML* ή μηνύματα ηλεκτρονικού ταχυδρομείου με επικεφαλίδες και περιεχόμενο σώματος. Εδώ υπεισέρχονται τεχνικές απόξεσης ιστού, οι οποίες επιτρέπουν σε εργαλεία να περιηγηθούν στις δομές ιστότοπων και να εξάγουν σχετικά σημεία δεδομένων. (TrustRadius, 2024)
- c. **Μη δομημένα δεδομένα:** Αυτή η κατηγορία περιλαμβάνει πληροφορίες χωρίς επίσημη δομή, όπως έγγραφα κειμένου, αναρτήσεις στα μέσα κοινωνικής δικτύωσης ή ηχογραφήσεις. Χρησιμοποιούνται τεχνικές επεξεργασίας φυσικής γλώσσας για την ανάλυση της γλώσσας και τον εντοπισμό βασικών οντοτήτων και σχέσεων εντός των δεδομένων. (Integrate.io., 2024)

Τα εργαλεία εξαγωγής πληροφοριών αυτοματοποιούν αυτήν την χρονοβόρα διαδικασία της χειροκίνητης συλλογής δεδομένων, βελτιώνοντας την αποδοτικότητα και την ακρίβεια. Τα εργαλεία εξαγωγής πληροφοριών βρίσκουν εφαρμογή σε διάφορες κλάδους όπως είναι :

- Χρηματοοικονομικά: Εξαγωγή χρηματοοικονομικών δεδομένων από αναφορές, τιμολόγια και συναλλαγές για τη διευκόλυνση της χρηματοοικονομικής ανάλυσης και της λήψης αποφάσεων. (Microblink, 2024)
- Υγειονομική περίθαλψη: Επεξεργασία ιατρικών αρχείων, δεδομένων απαιτήσεων και πληροφοριών κλινικών δοκιμών για την υποστήριξη της έρευνας, τη βελτίωση της φροντίδας των ασθενών και τη διασφάλιση της κανονιστικής συμμόρφωσης. (Rivery, 2024)
- Ηλεκτρονικό εμπόριο: Συγκέντρωση πληροφοριών για τα προϊόντα, κριτικές πελατών και δεδομένα τιμολόγησης από ιστότοπους ανταγωνιστών για ανάλυση της αγοράς και στρατηγικές τιμολόγησης.

Στην ουσία, τα εργαλεία εξαγωγής πληροφοριών δίνουν τη δυνατότητα στους οργανισμούς να αξιοποιήσουν τον τεράστιο όγκο των διαθέσιμων δεδομένων και να τα μετατρέψουν σε αξιοποιήσιμες πληροφορίες.

Τα πιο δημοφιλή εργαλεία για την εξαγωγή πληροφοριών καλύπτουν ποικίλες ανάγκες και πηγές δεδομένων. Αυτά τα εργαλεία είναι το :

1. *Google Cloud Document AI*: Αυτό το εργαλείο που βασίζεται στο cloud αξιοποιεί τη μηχανική μάθηση για την εξαγωγή δεδομένων από έγγραφα όπως τιμολόγια, συμβάσεις και ιατρικά αρχεία, καθιστώντας πολύτιμο εργαλείο σε διάφορους κλάδους. (Google Cloud., n.d.).
2. *Hevo Data*: Αυτή η πλατφόρμα προσφέρει μια ολοκληρωμένη λύση για την εξαγωγή δεδομένων, υποστηρίζοντας συνδέσεις με περισσότερες από 150 πηγές δεδομένων και παρέχοντας χαρακτηριστικά για μετασχηματισμό και εξαγωγή δεδομένων . (Hevo Data, n.d.)
3. *UiPath*: Αυτό το εργαλείο αυτοματοποίησης ρομποτικών διαδικασιών (*RPA*) μπορεί να ενσωματωθεί με δυνατότητες εξαγωγής πληροφοριών για την αυτοματοποίηση επαναλαμβανόμενων εργασιών που περιλαμβάνουν συλλογή και επεξεργασία δεδομένων. (UiPath, n.d.)

Τα εργαλεία αυτά αντιπροσωπεύουν ένα μικρό μέρος των διαθέσιμων λύσεων εξαγωγής πληροφοριών. Η επιλογή του κατάλληλου εργαλείου εξαρτάται από τις συγκεκριμένες απαιτήσεις, τις πηγές δεδομένων και τις επιθυμητές λειτουργίες.

2.1.3. Τα πλεονεκτήματα της εξαγωγής πληροφορίας

Η εξαγωγή πληροφοριών προσφέρει μια πληθώρα πλεονεκτημάτων σε διάφορους κλάδους και εφαρμογές. Αυτοματοποιώντας τη διαδικασία εξαγωγής συγκεκριμένων σημείων δεδομένων από διάφορες πηγές, η *IE* δίνει τη δυνατότητα στις επιχειρήσεις και τους οργανισμούς να ξεκλειδώσουν το πραγματικό δυναμικό των πληροφοριών τους. Τα οφέλη που αποκτούνται από την εξαγωγή πληροφορίας είναι τα εξής:

- ✓ **Ενισχυμένη αποδοτικότητα και ακρίβεια:** Ένα από τα σημαντικότερα πλεονεκτήματα της εξαγωγής πληροφοριών έγκειται στην ικανότητά της να αυτοματοποιεί τη διαδικασία εξαγωγής δεδομένων. Παραδοσιακά, η χειροκίνητη συλλογή δεδομένων από διάφορες πηγές μπορεί να είναι χρονοβόρα, επιρρεπής σε σφάλματα και να απαιτεί σημαντικούς ανθρώπινους πόρους. Τα εργαλεία *IE* αντιμετωπίζουν αυτή την πρόκληση αυτοματοποιώντας τη διαδικασία, μειώνοντας σημαντικά τον χρόνο και την προσπάθεια που απαιτούνται για τη συλλογή των επιθυμητών πληροφοριών. Αυτό μεταφράζεται σε βελτιωμένη αποδοτικότητα και επιτρέπει στο προσωπικό να επικεντρωθεί σε εργασίες υψηλότερης αξίας που απαιτούν ανθρώπινη κρίση και ανάλυση. (Microblink, 2024) Επιπλέον, η αυτοματοποίηση ελαχιστοποιεί τον κίνδυνο ανθρώπινου σφάλματος που συνδέεται με τη χειροκίνητη εισαγωγή δεδομένων. Τα εργαλεία *IE* αξιοποιούν προηγμένους αλγορίθμους και τεχνικές για να εξασφαλίσουν μεγαλύτερη ακρίβεια στα εξαγόμενα δεδομένα, οδηγώντας σε πιο αξιόπιστες πληροφορίες για περαιτέρω ανάλυση και λήψη αποφάσεων. (Integrate.io., 2024)
- ✓ **Βελτιωμένη προσβασιμότητα και ανάλυση δεδομένων:** Η εξαγωγή πληροφοριών διευκολύνει τον μετασχηματισμό δεδομένων από διάφορες πηγές σε δομημένη και ενοποιημένη μορφή. Αυτό επιτρέπει στους οργανισμούς να ενσωματώνουν απρόσκοπτα δεδομένα από διαφορετικές πηγές, όπως τα μέσα κοινωνικής δικτύωσης, τις αξιολογήσεις πελατών, τις οικονομικές εκθέσεις και τις ιστοσελίδες. Με τη δημιουργία ενός κεντρικού αποθετηρίου δομημένων δεδομένων, τα εργαλεία *IE* βελτιώνουν την προσβασιμότητα των δεδομένων και δίνουν στους χρήστες τη δυνατότητα να διεξάγουν ολοκληρωμένες αναλύσεις. Αυτό επιτρέπει τον εντοπισμό τάσεων, μοτίβων και πολύτιμων πληροφοριών που διαφορετικά θα μπορούσαν να παραμείνουν κρυμμένες μέσα στον τεράστιο όγκο των ακατέργαστων δεδομένων. Επιπροσθέτως, η δομημένη μορφή διευκολύνει τη χρήση διαφόρων εργαλείων και

τεχνικών ανάλυσης δεδομένων. Στατιστικό λογισμικό, αλγόριθμοι μηχανικής μάθησης και εργαλεία οπτικοποίησης δεδομένων μπορούν να εφαρμοστούν αποτελεσματικά στα εξαγόμενα δεδομένα, επιτρέποντας τη βαθύτερη εξερεύνηση και την αποκάλυψη κρυμμένων σχέσεων εντός των πληροφοριών. (Astera Software, 2023)

- ✓ Βελτιστοποιημένη λήψη αποφάσεων και ανταγωνιστικό πλεονέκτημα: Οι γνώσεις που αντλούνται από την εξαγωγή πληροφοριών δίνουν τη δυνατότητα στους οργανισμούς να λαμβάνουν αποφάσεις βάσει δεδομένων με μεγαλύτερη αυτοπεποίθηση. Για παράδειγμα, οι επιχειρήσεις και οι οργανισμοί που βρίσκονται στο χρηματοπιστωτικό τομέα, η εξαγωγή πληροφοριών μπορεί να χρησιμοποιηθεί για την εξαγωγή οικονομικών δεδομένων από εκθέσεις και άρθρα ειδήσεων, διευκολύνοντας την τεκμηριωμένη λήψη επενδυτικών αποφάσεων. Αξιοποιώντας τη δύναμη των εξαγόμενων δεδομένων, οι οργανισμοί αποκτούν ανταγωνιστικό πλεονέκτημα στους αντίστοιχους τομείς τους. (Rivery, 2024)

Πέρα από τα βασικά οφέλη που αναφέρθηκαν παραπάνω, η εξαγωγή πληροφοριών προσφέρει διάφορα πρόσθετα πλεονεκτήματα όπως είναι :

- ✓ Μειωμένο κόστος: Η αυτοματοποίηση των διαδικασιών εξαγωγής δεδομένων οδηγεί σε σημαντική εξοικονόμηση κόστους που συνδέεται με τη χειροκίνητη συλλογή και ανάλυση δεδομένων.
- ✓ Βελτιωμένη κανονιστική συμμόρφωση: Ορισμένοι κλάδοι έχουν αυστηρούς κανονισμούς σχετικά με τη συλλογή και την υποβολή εκθέσεων δεδομένων. Τα δεδομένα που εξάγονται μπορούν εύκολα να μορφοποιηθούν ώστε να συμμορφώνονται με αυτούς τους κανονισμούς.
- ✓ Βελτιωμένη εξυπηρέτηση πελατών: Οι επιχειρήσεις μπορούν να αξιοποιήσουν τα εξαγόμενα δεδομένα πελατών για να εξατομικεύσουν τις προσφορές τους και να παρέχουν μια πιο ευέλικτη εμπειρία εξυπηρέτησης πελατών.

Συμπερασματικά, η εξαγωγή πληροφοριών αποτελεί ένα ισχυρό εργαλείο για τους οργανισμούς που επιδιώκουν να ξεκλειδώσουν την πραγματική αξία των δεδομένων τους. Με την αυτοματοποίηση της εξαγωγής δεδομένων, τη βελτίωση της ακρίβειας και τη διευκόλυνση της ανάλυσης δεδομένων, η *IE* δίνει τη δυνατότητα στις επιχειρήσεις να λαμβάνουν

τεκμηριωμένες αποφάσεις, να αποκτούν ανταγωνιστικό πλεονέκτημα και να επιτυγχάνουν τους στρατηγικούς τους στόχους.

2.2. Τεχνικές εξαγωγής πληροφορίας από διαδικτυακές αγγελίες εργασίας

Ο συνεχώς αυξανόμενος όγκος των αγγελιών εργασίας στο διαδίκτυο αποτελεί πρόκληση τόσο για τους αναζητούντες εργασία όσο και για τους ερευνητές. Το ξεδιάλεγμα αμέτρητων αγγελιών για την εύρεση σχετικών πληροφοριών μπορεί να είναι μια χρονοβόρα και κουραστική εργασία. Ευτυχώς, οι εξελίξεις στις τεχνικές εξόρυξης κειμένου προσφέρουν πολλές υποσχόμενες δυνατότητες αυτοματοποιημένης εξαγωγής πληροφοριών, καθιστώντας τη διαδικασία πιο αποτελεσματική και διορατική. Οι τεχνικές που χρησιμοποιούνται για την εξαγωγή πολύτιμων δεδομένων από διαδικτυακές αγγελίες εργασίας είναι οι εξής:

1. Εξαγωγή και αντιστοίχιση λέξεων-κλειδιών: Μια θεμελιώδης προσέγγιση περιλαμβάνει τον εντοπισμό λέξεων-κλειδιών που αντιπροσωπεύουν ουσιώδεις πληροφορίες, όπως τίτλους θέσεων εργασίας, δεξιότητες και απαιτήσεις εμπειρίας. Αυτό μπορεί να επιτευχθεί μέσω απλών τεχνικών όπως οι κανονικές εκφράσεις ή πιο εξελιγμένων μεθόδων όπως η αναγνώριση ονομαστικών οντοτήτων (*Named Entity Recognition - NER*) (Nadeau, 2007). Οι αλγόριθμοι *NER* αξιοποιούν μοντέλα μηχανικής μάθησης για τον εντοπισμό και την ταξινόμηση συγκεκριμένων οντοτήτων μέσα στο κείμενο, όπως τοποθεσίες, οργανισμοί και, στην προκειμένη περίπτωση, όροι που σχετίζονται με την εργασία.
2. Επισήμανση μέρους του λόγου :Η τεχνική αυτή αποδίδει γραμματικές ετικέτες (π.χ. ουσιαστικό, ρήμα, επίθετο) σε κάθε λέξη μιας πρότασης. Οι πληροφορίες αυτές βοηθούν στον εντοπισμό του πλαισίου και των σχέσεων μεταξύ των λέξεων. Για παράδειγμα, η φράση "πρέπει να έχει ισχυρές επικοινωνιακές δεξιότητες" μπορεί να αναλυθεί αναγνωρίζοντας την "επικοινωνία" ως ουσιαστικό και την "ισχυρή" ως επίθετο, υποδεικνύοντας μια επιθυμητή δεξιότητα για τη θέση. (Bird, 2009)
3. Κανονικές εκφράσεις: Οι κανονικές εκφράσεις είναι ισχυρά εργαλεία για την αντιστοίχιση μοτίβων μέσα στο κείμενο. Επιτρέπουν τον ορισμό συγκεκριμένων

μοτίβων για τον εντοπισμό των επιθυμητών πληροφοριών. Για παράδειγμα, μια κανονική έκφραση όπως "έτη εμπειρίας" μπορεί να χρησιμοποιηθεί για την εξαγωγή του αριθμού των ετών εμπειρίας που αναφέρονται στην αγγελία εργασίας. (Friedl, 2006)

4. Τεχνικές μηχανικής μάθησης :Οι επιβλεπόμενοι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν σε επισημασμένα σύνολα δεδομένων από αγγελίες εργασίας, όπου συγκεκριμένες πληροφορίες που έχουν ήδη εντοπιστεί. Οι αλγόριθμοι αυτοί στη συνέχεια μαθαίνουν να αναγνωρίζουν μοτίβα και να εξάγουν παρόμοιες πληροφορίες από νέες αθέατες αγγελίες εργασίας. Τεχνικές όπως οι μηχανές διανυσμάτων υποστήριξης (*SVM*) και τα τυχαία δάση έχουν δείξει πολλά υποσχόμενα αποτελέσματα στην ακριβή ταξινόμηση περιγραφών θέσεων εργασίας και στην εξαγωγή σχετικών λεπτομερειών. (James, 2013)

5. Τεχνικές επεξεργασίας φυσικής γλώσσας: Η *NLP* περιλαμβάνει ένα ευρύτερο φάσμα τεχνικών που υπερβαίνουν την απλή αντιστοίχιση λέξεων-κλειδιών και εμβαθύνουν στη σημασιολογική σημασία του κειμένου. Τεχνικές όπως η ανάλυση συναισθήματος μπορούν να χρησιμοποιηθούν για τη μέτρηση του συνολικού τόνου και της εταιρικής κουλτούρας που αντικατοπτρίζεται στην περιγραφή της θέσης εργασίας. Επιπλέον, η μοντελοποίηση θεμάτων μπορεί να εντοπίσει τα υποκείμενα θέματα και τα απαιτούμενα σύνολα δεξιοτήτων για την θέση εργασίας που ζητείται. (Bird, 2009)

Οι τεχνικές αυτές προσφέρουν διάφορες προσεγγίσεις που χρησιμοποιούνται για την εξαγωγή πληροφοριών από διαδικτυακές αγγελίες εργασίας καθώς, ο τομέας της εξόρυξης κειμένου συνεχίζει να εξελίσσεται σε βάθος χρόνου.

2.2.1. Η χρήση των APIs των ιστότοπων αγγελιών εργασίας

Στη σημερινή ψηφιακή εποχή, τα *APIs* για την αναζήτηση εργασίας έχουν γίνει απαραίτητο εργαλείο τόσο για τους αναζητούντες εργασία όσο και για τους προσλαμβάνοντες. Τα *APIs* αυτά παρέχουν πρόσβαση σε καταχωρίσεις θέσεων εργασίας από διάφορες πηγές, διευκολύνοντας την εύρεση της κατάλληλης θέσης εργασίας ή του κατάλληλου υποψηφίου.

Τα *APIs* αναζήτησης εργασίας αποτελούν διεπαφές προγραμματισμού εφαρμογών που επιτρέπουν στους προγραμματιστές να έχουν πρόσβαση σε λίστες θέσεων εργασίας και συναφή δεδομένα από διάφορους ιστότοπους αναζήτησης εργασίας και πίνακες εργασίας. Επιπροσθέτως, παρέχουν έναν τυποποιημένο τρόπο για τους προγραμματιστές να ενσωματώνουν λειτουργίες αναζήτησης εργασίας στις εφαρμογές, τους ιστότοπους ή τις υπηρεσίες τους. Τις περισσότερες φορές παρέχουν πρόσβαση σε λίστες θέσεων εργασίας, περιγραφές θέσεων εργασίας, πληροφορίες για την εταιρεία, δεδομένα τοποθεσίας και άλλες σχετικές πληροφορίες που σχετίζονται με την αναζήτηση εργασίας. Μπορούν να χρησιμοποιηθούν για τη δημιουργία μηχανών αναζήτησης θέσεων εργασίας, πλατφορμών αντιστοίχισης θέσεων εργασίας, εργαλείων ανάπτυξης καριέρας και άλλων εφαρμογών που σχετίζονται με την εργασία. Τα *APIs* αναζήτησης εργασίας μπορούν να βοηθήσουν τους αναζητούντες εργασία να βρουν σχετικές ευκαιρίες απασχόλησης και τους εργοδότες να προσεγγίσουν μια ευρύτερη δεξαμενή υποψηφίων. Έτσι, βοηθήσουν τους υπεύθυνους προσλήψεων και τους επαγγελματίες ανθρώπινου δυναμικού να βελτιώσουν τις διαδικασίες πρόσληψης και να βελτιώσουν τις στρατηγικές πρόσληψης. Παρακάτω αναφέρονται οι περιπτώσεις χρήσης των *APIs* αναζήτησης εργασίας:

- i. Δικτυακοί τόποι συγκέντρωσης θέσεων εργασίας: Αφορά την συλλογή των καταχωρημένων θέσεων εργασίας από πολλαπλές πηγές και την εμφάνιση στην πλατφόρμα τους. Αυτό επιτρέπει στους αναζητούντες εργασία να αναζητούν θέσεις εργασίας από διάφορες πηγές.
- ii. Υπηρεσίες καριέρας: Οι υπηρεσίες σταδιοδρομίας, όπως οι υπηρεσίες δημιουργίας βιογραφικών σημειωμάτων, τα εργαλεία προετοιμασίας για συνεντεύξεις και οι υπηρεσίες συμβουλευτικής σταδιοδρομίας, μπορούν να χρησιμοποιούν *APIs* αναζήτησης θέσεων εργασίας για να παρέχουν καταχωρίσεις θέσεων εργασίας στους χρήστες τους. Αυτό βοηθά τους αναζητούντες εργασία να βρίσκουν σχετικές ευκαιρίες απασχόλησης και να υποβάλλουν αίτηση απευθείας από την πλατφόρμα υπηρεσιών καριέρας.
- iii. Πρακτορεία προσλήψεων: Τα γραφεία προσλήψεων μπορούν να χρησιμοποιούν *APIs* αναζήτησης θέσεων εργασίας για να βρουν καταχωρημένες θέσεων εργασίας που ταιριάζουν με τις δεξιότητες και τα προσόντα των πελατών τους. Αυτό τους επιτρέπει

να παρέχουν στοχευμένες ευκαιρίες απασχόλησης στους πελάτες τους και να αυξάνουν τις πιθανότητες εύρεσης εργασίας.

- iv. Λογισμικό ανθρώπινου δυναμικού: Το λογισμικό ανθρώπινου δυναμικού μπορεί να χρησιμοποιήσει *APIs* αναζήτησης θέσεων εργασίας για να ενσωματώσουν καταχωρημένες θέσεων εργασίας στην πλατφόρμα και να βελτιώσει στη διαδικασία πρόσληψης. Αυτό επιτρέπει στους επαγγελματίες του τομέα ανθρώπινου δυναμικού να διαχειρίζονται τις θέσεις εργασίας, να παρακολουθούν τους αιτούντες και να επικοινωνούν με τους υποψηφίους.
- v. Έρευνα και ανάλυση: Οι ερευνητές και οι αναλυτές μπορούν να χρησιμοποιούν *APIs* αναζήτησης εργασίας για να συλλέγουν δεδομένα σχετικά με τις τάσεις της απασχόλησης, τους μισθούς και τη ζήτηση θέσεων εργασίας. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για την ενημέρωση πολιτικών αποφάσεων, κλαδικών εκθέσεων και ακαδημαϊκής έρευνας. (Kelly, 2023)

Ο συνεχώς αυξανόμενος όγκος των διαδικτυακών αγγελιών εργασίας αποτελεί πολύτιμες πληροφορίες τόσο για όσους αναζητούν εργασία όσο και για ερευνητές. Ωστόσο, η πλοήγηση σε αυτό το τεράστιο τοπίο για την εύρεση σχετικών πληροφοριών μπορεί να είναι μια χρονοβόρα και επίπονη εργασία. Για τον λόγο αυτό, η εξέλιξη της τεχνολογίας έχει συμβάλει στον προγραμματισμό εφαρμογών που προσφέρονται από ιστότοπους ανάρτησης θέσεων εργασίας με αποτέλεσμα να παρέχουν μια ισχυρή λύση για βελτιωμένη πρόσβαση και ανάλυση δεδομένων. Τα πιο δημοφιλή *APIs* αναζήτησης εργασίας είναι τα εξής:

- *Indeed API*: Προσφέρει εκτεταμένη κάλυψη στις αγγελίες εργασίας από διάφορους κλάδους και τοποθεσίες. Καθώς, δίνει προηγμένες δυνατότητες αναζήτησης επιτρέποντας το φιλτράρισμα βάσει συγκεκριμένων κριτηρίων, όπως ο τύπος εργασίας, το μέγεθος της εταιρείας και το εύρος του μισθού. (Indeed, 2024)
- *LinkedIn Talent Solutions*: Παρέχει μια στοχευμένη πρόσβαση καθιστώντας ιδανικό για προσλήψεις και άτομα που αναζητούν συγκεκριμένες επαγγελματικές διαδρομές. Δίνει λεπτομερείς σε πληροφορίες σχετικά με τις περιγραφές θέσεων εργασίας, τα προφίλ εταιρειών και τις δεξιότητες των υποψηφίων. Επιπλέον, ενσωματώνει το

δικτύου ώστε να αξιοποιεί τη δύναμη του επαγγελματικού δικτύου του *LinkedIn* για πιθανές συνδέσεις. (inDevelopers, 2024)

- *Glassdoor API*: Το περιεχόμενο δημιουργείται από χρήστες προσφέροντας έτσι μια ευρύτερη προοπτική για τις εταιρείες και τις θέσεις εργασίας. Συνεισφέρει προσαρμόσιμη αναζήτηση επιτρέποντας το φιλτράρισμα με βάση την αξιολόγηση της εταιρείας, τη δυσκολία της συνέντευξης και τα προσφερόμενα οφέλη. (Glassdoor API, 2024)

Ορισμένα *APIs* προσφέρουν δωρεάν βαθμίδες, άλλα έχουν όρια χρήσης ή απαιτούν συνδρομές επί πληρωμή για εκτεταμένη πρόσβαση. Η κάλυψη των δεδομένων αποτελεί ένα σημαντικό ζήτημα καθώς, η πληρότητα των καταλόγων θέσεων εργασίας μπορεί να διαφέρει ανάλογα με τις συνεργασίες και τους τομείς εστίασης του παρόχου. Αξίζει να σημειωθεί, ότι η επιλογή του σωστού *APIs* γίνεται με την αξιολόγηση των αναγκών καθορίζοντας με σαφήνεια το κοινό-στόχο, τα επιθυμητά σημεία δεδομένων και τους περιορισμούς του προϋπολογισμού, τις λειτουργίες που προσφέρονται τα διαφορετικά *APIs*.

2.2.2. Web Scrapping

Η εύρεση δεδομένων για βιομηχανικά και ακαδημαϊκά ερευνητικά έργα αποτελούσε μια δύσκολη και δαπανηρή διαδικασία. Σήμερα, η ψηφιοποίηση και η εικονικοποίηση των κοινωνικών διεργασιών έχουν ως αποτέλεσμα να υπάρχουν δεδομένων διαθέσιμα στον ιστό. Τα δεδομένα αυτά παρέχουν μια λεπτομερή, σε πραγματικό χρόνο αναπαράσταση πολυάριθμων διαδικασιών, σχέσεων και αλληλεπιδράσεων στον κοινωνικό-υλικό χώρο. Έτσι, αυτοί οι τεράστιοι όγκοι δεδομένων του μεγάλου ιστού προσφέρουν στους ακαδημαϊκούς ερευνητές άφθονες ευκαιρίες για την απάντηση νέων και παλαιών ερευνητικών ερωτημάτων με μεγαλύτερη αυστηρότητα, ακρίβεια και χρονικά περιθώρια. Οι επαγγελματίες μπορούν να αξιοποιήσουν αυτά τα δεδομένα για την ανάπτυξη καλύτερης κατανόησης των πελατών τους, τη διαμόρφωση στρατηγικών με βάση αυτά τα ευρήματα με στόχο την βελτίωση των οργανωτικών επιδόσεων. Παρόλα αυτά, η αξιοποίηση αυτών των τεράστιων όγκων δεδομένων του παγκόσμιου ιστού παρουσιάζει σοβαρές τεχνικές, νομικές και ηθικές προκλήσεις. Ενώ, έχουν πολλαπλασιαστεί τα εργαλεία και οι τεχνολογίες που μπορούν να χρησιμοποιηθούν για το *Web Scrapping*. Όμως, η νομιμότητα και η ηθική της συλλογής δεδομένων από τον παγκόσμιο ιστό εξακολουθούν να είναι μείζον πρόβλημα. Ενώ, τα υφιστάμενα νομικά πλαίσια μπορούν να εφαρμοστούν, σε κάποιο βαθμό, στην αναδυόμενη

πρακτική του *Web Scraping*, τα ηθικά ζητήματα που περιβάλλουν το *Web Scraping* έχουν αγνοηθεί σε μεγάλο βαθμό.

Τα δεδομένα που είναι διαθέσιμα στον παγκόσμιο ιστό αποτελούνται από δομημένα, ημιδομημένα και μη δομημένα ποσοτικά και ποιοτικά δεδομένα με τη μορφή ιστοσελίδων, πινάκων *HTML*, βάσεων δεδομένων στον παγκόσμιο ιστό, μηνυμάτων ηλεκτρονικού ταχυδρομείου, *tweets*, αναρτήσεων σε ιστολογία, φωτογραφιών, βίντεο κ.λπ. Η αξιοποίηση των δεδομένων του μεγάλου ιστού απαιτεί την αντιμετώπιση ορισμένων τεχνικών ζητημάτων που σχετίζονται με τον όγκο, την ποικιλία, την ταχύτητα και την ελικρίνεια των δεδομένων στον ιστό. (Krotov, 2020)

Το *Web Scraping* είναι μια τεχνική εξαγωγής δεδομένων από ιστότοπους με τη χρήση αυτοματοποιημένων σεναρίων ή προγραμμάτων. Οι πληροφορίες αυτές συλλέγονται και στη συνέχεια εξάγονται σε μορφή που είναι πιο χρήσιμη για τον χρήστη. Είτε, πρόκειται για ένα υπολογιστικό φύλλο είτε για ένα *API*. Το *Web Scraping* μπορεί να γίνει χειροκίνητα, στις περισσότερες περιπτώσεις όμως προτιμότερη είναι η χρήση αυτοματοποιημένων εργαλείων. Καθώς, μπορεί να είναι λιγότερο δαπανηρά και να λειτουργούν με ταχύτερο ρυθμό. (Perez, 2023) Το *Web Scraping* είναι μια τεχνική που είναι χρήσιμη για έργα επιστήμης δεδομένων που απαιτούν μεγάλα και ποικίλα σύνολα δεδομένων, όπως η ανάλυση τάσεων της αγοράς εργασίας, η ζήτηση δεξιοτήτων, τα εύρη μισθών κ.λπ. Οι πίνακες ευρέσεως εργασίας αποτελούν μια εξαιρετική πηγή δεδομένων που μπορούν να παράσχουν πληροφορίες για διάφορες πτυχές της αγοράς εργασίας. Για παράδειγμα ο αριθμός και η κατανομή των κενών θέσεων εργασίας, οι πιο κοινές και επιθυμητές δεξιότητες και προσόντα, οι μισθοί και οι παροχές που προσφέρονται από τους εργοδότες και το επίπεδο ανταγωνισμού για διάφορους τύπους θέσεων εργασίας. Με τη σάρωση των πινάκων ευρέσεως εργασίας, μπορούμε να έχουμε μια εικόνα σχετικά με τις καλύτερες πόλεις ή χώρες για να βρούμε θέσεις εργασίας στον τομέα που σας ενδιαφέρει, ποιες είναι οι δεξιότητες και τα προσόντα που αναζητούν οι εργοδότες στον τομέα σας. Επιπλέον, μπορούμε να μάθουμε όσο αναφορά για ποσά που μπορεί να κερδίσετε στον επιθυμητό ρόλο, στην τοποθεσία που επιθυμείτε, πώς να βελτιστοποιήσετε το βιογραφικό σας σημείωμα και τη συνοδευτική σας επιστολή ώστε να ταιριάζουν με τις απαιτήσεις και τις λέξεις-κλειδιά της θέσης εργασίας. Το *Web Scraping* πινάκων εργασίας απαιτεί την τήρηση ορισμένων γενικών βημάτων:

- a. Προσδιορισμός του ιστότοπου-στόχου και των δεδομένων που θέλουμε να εξάγουμε.
- b. Επιθεώρηση της δομής του ιστότοπου και των στοιχείων *HTML*.

- c. Συγγραφή ενός σεναρίου ή προγράμματος για την αποστολή αιτημάτων και την ανάλυση της απάντησης *HTML* και αποθήκευση των εξαγόμενων δεδομένων σε κατάλληλη μορφή.

Υπάρχουν πολλά διαθέσιμα εργαλεία και βιβλιοθήκες που βοηθούν *Web Scrapping* ιστοσελίδων, ανάλογα με το επίπεδο εμπειρίας και τις προτιμήσεις του χρήστη. Μερικά από αυτά είναι:

- Το *Beautiful Soup*, μια βιβλιοθήκη *Python*, που μπορεί να αναλύσει έγγραφα *HTML* και *XML* και να εξάγει δεδομένα χρησιμοποιώντας διάφορες μεθόδους και φίλτρα.
- Το *Scrapy*, είναι ένα δωρεάν και ανοιχτού κώδικα πλαίσιο λογισμικού για ανίχνευση ιστοσελίδων σε *Python* είναι κατάλληλη για σύνθετα και μεγάλης κλίμακας έργα *Web scrapping* ιστού.
- Το *Selenium* είναι ένα εργαλείο που αυτοματοποιεί ενέργειες του προγράμματος περιήγησης ιστού. Επίσης, περιέχει δυναμικό και διαδραστικό περιεχόμενο.
- Το *Requests* είναι μια βιβλιοθήκη *Python* που στέλνει αιτήσεις *HTTP* σε ιστότοπους και αποκτά πρόσβαση στα δεδομένα και τις επικεφαλίδες της απάντησης.
- Το *Pandas* είναι μια βιβλιοθήκη *Python* που χειρίζεται και αναλύει δομές δεδομένων, όπως πλαίσια δεδομένων και σειρές μπορεί επίσης, να διαβάζει και να γράφει δεδομένα σε διάφορες μορφές. (Linkedin, 2024)

Το *Web Scrapping* αποτελεί μια διαδικασία εξαγωγής δεδομένων από ιστότοπους, προσφέρει μια σειρά από πλεονεκτήματα, αλλά έχει και εγγενή μειονεκτήματα όπως, η κατανόηση για την υπεύθυνη και ηθική συλλογή δεδομένων. Τα οφέλη που μπορούμε να αποκομίσουμε είναι τα κάτωθι:

- 🚦 Οικονομικά αποδοτικά: Σε σύγκριση με τη χειροκίνητη συλλογή δεδομένων, το *Web Scrapping* αυτοματοποιεί τη διαδικασία, μειώνοντας σημαντικά τον χρόνο και την επένδυση πόρων.
- 🚦 Επεκτασιμότητα: Μπορούν να εξαχθούν αποτελεσματικά μεγάλοι όγκοι δεδομένων, γεγονός που την καθιστά κατάλληλη για εργασίες όπως η έρευνα αγοράς ή ανάλυση συναισθήματος.
- 🚦 Διαθεσιμότητα δεδομένων: Το *Web Scrapping* ξεκλειδώνει την πρόσβαση σε πολύτιμα δεδομένα που δεν είναι άμεσα διαθέσιμα μέσω των παραδοσιακών μεθόδων.

- ✚ Δεδομένα σε πραγματικό χρόνο: Το *Web Scrapping* μπορούν να συλλάβουν δυναμικά δεδομένα που ενημερώνονται συχνά, παρέχοντας πολύτιμες πληροφορίες για τις τάσεις.
- ✚ Ανάλυση δεδομένων: Τα δεδομένα που εξάγονται μπορούν να ενσωματωθούν σε διάφορα αναλυτικά εργαλεία για περαιτέρω εξερεύνηση και ανακάλυψη προτύπων. (Ghani, 2017)

Από την άλλη πλευρά εκτός από τα πλεονεκτήματα που μπορεί να μας προσφέρει το *Web Scrapping* υπάρχουν και μειονεκτήματα που μπορούν να επηρεάσουν αρνητικά την επιλογή της διαδικασίας *Web Scrapping*. Όπως είναι :

- ✚ Ηθικές ανησυχίες: Το *Web Scrapping* χωρίς άδεια μπορεί να παραβιάσει τους όρους παροχής υπηρεσιών του ιστότοπου και ενδεχομένως να είναι παράνομη.
- ✚ Ποιότητα δεδομένων: Τα δεδομένα που εξάγονται ενδέχεται να είναι ασυνεπής, ανακριβής ή ελλιπής, απαιτώντας πρόσθετο καθαρισμό και επεξεργασία.
- ✚ Σταθερότητα ιστότοπου: Το συχνό *Web Scrapping* μπορεί να υπερφορτώσει τους διακομιστές ιστότοπων, οδηγώντας σε διακοπές ή αποκλεισμό *IP*.
- ✚ Νομικά ζητήματα: Το *Web Scrapping* περιεχομένου που προστατεύεται από πνευματικά δικαιώματα ή η παραβίαση των κατευθυντήριων γραμμών robots.txt μπορεί να οδηγήσει σε νομικές επιπτώσεις. (Hjorth, 2016)

Το *Web Scrapping* ιστοσελίδων έχει τεράστιες δυνατότητες για διάφορες εφαρμογές. Ωστόσο, απαιτείται προσοχή και προτεραιότητα σε δεοντολογικές πρακτικές. Η κατανόηση των νομικών επιπτώσεων, ο σεβασμός των κατευθυντήριων γραμμών του ιστότοπου και η διασφάλιση του απορρήτου των δεδομένων είναι απαραίτητα για την υπεύθυνη συλλογή δεδομένων μέσω του *Web Scrapping*.

2.2.3. Η χρήση των RSS Feeds

Το διαδίκτυο έχει αλλάξει δραματικά την ποσότητα και τους τύπους των πληροφοριών που μπορούν να παρουσιαστούν και να μοιραστούν. Λειτουργεί ως πλατφόρμα για την απόκτηση και διανομή πληροφοριών σε χρήστες σε όλο τον κόσμο. Η διαχείριση όλων αυτών των πηγών πληροφόρησης μπορεί να αποτελέσει μια δύσκολη διαδικασία. Εδώ και πολλές

δεκαετίες, η εξέλιξη της τεχνολογίας έχει εισάγει τις *RSS* συχνά αναφέρονται ως “*Really Simple Syndication*”, (Πολύ Απλή Διανομή) ή παλαιότερα τον όρο *RDF Site Summary* (Σύνοψη Σελίδας *RDF*), αφορά σε μία μέθοδο διανομής ψηφιακού πληροφοριακού περιεχομένου μέσα από τον παγκόσμιο ιστό, στηριγμένη στη γλώσσα σήμανσης *XML*. Ένας χρήστης του Διαδικτύου μπορεί να ενημερώνεται αυτόματα για το ανανεωμένο περιεχόμενο από όσους ιστότοπους παρέχουν ενημερώσεις μέσω *RSS* αρκεί να έχει εγγραφεί ως συνδρομητής σε αυτές. Οι ροές *RSS* συνήθως περιέχουν τον τίτλο επιμέρους ιστοσελίδων, σύνοψη του περιεχομένου τους, σχετικά μεταδεδομένα, ημερομηνία δημοσίευσης κλπ. Ο χρήστης χρειάζεται ένα ειδικό λογισμικό «αναγνώστη» («*feed aggregator*» ή «*feed reader*») το οποίο σε κάποιες περιπτώσεις μπορεί να είναι ενσωματωμένο στον περιηγητή Ιστού του χρήστη. Ο χρήστης εισάγει στον αναγνώστη στη διεύθυνση *URL* της ζητούμενης ροής, ώστε να εγγραφεί σε αυτήν ως συνδρομητής. Ο αναγνώστης ελέγχει τακτικά τις καταγεγραμμένες ροές για ενημερώσεις περιεχομένου. Επίσης, παρέχει στον εκάστοτε χρήστη μία κοινή και εξατομικευμένη διασύνδεση για όλες τις διαφορετικές πηγές στις οποίες είναι εγγεγραμμένος ο συνδρομητής. (Wikipedia, 2023) Οι τρεις πιο δημοφιλείς αναγνώστες ροών *RSS* είναι το :

- i. *Feedly*: Αποτελεί μια δημοφιλής πλατφόρμα ανάγνωσης ροών *RSS* και συγκέντρωσης περιεχομένου που επιτρέπει στους χρήστες να εγγραφούν και να οργανώσουν τους αγαπημένους τους ιστότοπους, ιστολογία, πηγές ειδήσεων και άλλες διαδικτυακές πηγές περιεχομένου σε ένα μέρος. Προσφέρει μια βελτιωμένη και προσαρμοσμένη διεπαφή για την διαχείριση ροών *RSS* σε διάφορες συσκευές και πλατφόρμες.
- ii. *NewsBlur*: Είναι ένα πρόγραμμα ανάγνωσης ροών *RSS* και συγκέντρωσης περιεχομένου που βασίζεται στον ιστό που έχει σχεδιαστεί για να βοηθά τους χρήστες να παραμένουν ενημερωμένοι για τα πιο πρόσφατα νέα και ενημερώσεις. Καθώς ,προσφέρει χαρακτηριστικά πλατφόρμας για την ανακάλυψη, την εγγραφή και την οργάνωση ροών *RSS*, καθώς και για την ανάγνωση και την κοινή χρήση περιεχομένου σε διάφορες συσκευές και πλατφόρμες.
- iii. *Feedbin*: Απαρτίζεται από ένα πρόγραμμα ανάγνωσης ροών *RSS* και υπηρεσία συγκέντρωσης περιεχομένου. Το *Feedbin* προσφέρει μια καθαρή διεπαφή με την ανάγνωση και ποικιλία λειτουργιών για την προσαρμογή εμπειρίας, όπως φιλτράρισμα ροών ανά κατηγορία και επιλογή προτιμώμενων γραμματοσειρών για άρθρα.

Τα *RSS Feeds* μπορούν να χρησιμοποιηθούν εξαιρετικά για την εξαγωγή πληροφοριών από διαδικτυακές αγγελίες εργασίας. Παρακάτω αναφέρονται κάποιοι τρόποι όπου μπορούμε να χρησιμοποιήσουμε τα *RSS Feeds*:

1. Εγγραφή σε *RSS Feeds* από ιστότοπους αγγελιών εργασίας: Οι περισσότεροι ιστότοποι αγγελιών εργασίας παρέχουν *RSS Feeds* για τις νέες αγγελίες. Μπορείτε να εγγραφείτε σε αυτά τα Feeds χρησιμοποιώντας έναν *RSS reader* ή κώδικα προγραμματισμού.
2. Αυτοματοποιημένη ανάκτηση και ανάλυση *Feeds*: Μπορείτε να χρησιμοποιήσετε ένα σενάριο προγραμματισμού για να ανακτήσετε αυτόματα τα *Feeds* από διάφορους ιστότοπους και να αναλύσετε τα δεδομένα για να εξάγετε σημαντικές πληροφορίες.
3. Φιλτράρισμα και κατηγοριοποίηση: Μετά τη λήψη των *Feeds*, μπορούμε να εφαρμόσουμε φίλτρα για να επιλέξετε μόνο τις αγγελίες που σας ενδιαφέρουν βάσει κριτηρίων όπως η τοποθεσία, η κατηγορία εργασίας, η εταιρεία κ.λπ.
4. Αυτόματη ειδοποίηση και αποθήκευση: Με τη χρήση ενός σεναρίου προγραμματισμού, μπορείτε να δημιουργήσετε μια αυτόματη διαδικασία που θα ειδοποιεί εσάς ή άλλους χρήστες για νέες αγγελίες εργασίας που πληρούν συγκεκριμένα κριτήρια και να αποθηκεύει τα δεδομένα σε μια βάση δεδομένων ή σε ένα αρχείο.
5. Ανάλυση δεδομένων και τάσεων: Αφού συλλέξετε δεδομένα από πολλές αγγελίες εργασίας μέσω των *Feeds*, μπορείτε να τα αναλύσετε για να ανακαλύψετε τάσεις στην αγορά εργασίας, όπως ζητούμενες δεξιότητες, μισθοί και τοποθεσίες εργασίας.

Με την χρήση των *RSS Feeds* και την αυτοματοποιημένη επεξεργασία των δεδομένων, μπορείτε να δημιουργήσετε έναν αποτελεσματικό τρόπο για την εξαγωγή πληροφοριών από διαδικτυακές αγγελίες εργασίας.

2.3. Προβλήματα που αντιμετωπίζονται στην ανάλυση και επεξεργασία πληροφοριών από διαδικτυακές αγγελιών εργασίας

Η εξέλιξη της τεχνολογίας και του διαδικτύου προσφέρει την δυνατότητα στους αναζητούντες εργασίας να μπορούν πλέον να έχουν πρόσβαση σε μια τεράστια πηγή ευκαιριών, ενώ οι εργοδότες μπορούν να προσεγγίσουν ένα ευρύτερο φάσμα δυνητικών υποψηφίων. Ωστόσο, η ανάλυση και η επεξεργασία των πληροφοριών από αυτές τις αγγελίες μπορεί να είναι ένα πολύπλοκη διαδικασίας με προκλήσεις. Η κατανόηση αυτών των εμποδίων είναι ζωτικής σημασίας τόσο για τους εργοδότες που επιδιώκουν να προσελκύσουν το κατάλληλο άτομο όσο και για τους αναζητούντες εργασία που ελπίζουν να εντοπίσουν την

καλύτερη δυνατή εφαρμογή των δεξιοτήτων τους. Ωστόσο, η εξαγωγή σημαντικών πληροφοριών από αυτές τις δημοσιεύσεις μπορεί να είναι μια ύπουλη προσπάθεια. Η ασυνεπής μορφοποίηση, η ασαφής γλώσσα και ο τεράστιος όγκος δεδομένων δημιουργούν σημαντικές προκλήσεις για τις παραδοσιακές μεθόδους επεξεργασίας πληροφοριών. Τα βασικά προβλήματα που αντιμετωπίζουμε κατά την ανάλυση και την επεξεργασία πληροφοριών από διαδικτυακές αγγελίες εργασίας:

- **Ασυνεπής μορφοποίηση και μη δομημένα δεδομένα:** Οι διαδικτυακές αναρτήσεις θέσεων εργασίας δεν διαθέτουν τυποποιημένη μορφή. Εμφανίζονται σε διάφορες πλατφόρμες με διαφορετικές διατάξεις, γραμματοσειρές και χρήση λέξεων-κλειδιών. Αυτή η ασυνέπεια δυσκολεύει τις παραδοσιακές μεθόδους ανάλυσης να εξαγάγουν με ακρίβεια τα σχετικά σημεία δεδομένων. Οι τίτλοι μπορεί να μην αντικατοπτρίζουν ξεκάθαρα τις επαγγελματικές ευθύνες. Επιπλέον, παρουσιάζεται ασυνέπεια σε σημαντικές πληροφορίες όπως το εύρος μισθών συχνά δεν βρίσκεται σε ευδιάκριτο σημείο μέσα στο κείμενο. (Zhi Hong, 2021)
- **Ποιότητα δεδομένων :** Οι διαδικτυακές αναρτήσεις εργασίας μπορεί να διαφέρουν σημαντικά ως προς την ποιότητα, την πληρότητα και την ακρίβεια του περιεχομένου. Ορισμένες αναρτήσεις μπορεί να περιέχουν ελλειπείς ή παραπλανητικές πληροφορίες, γεγονός που καθιστά δύσκολη την εξαγωγή σημαντικών πληροφοριών.
- **Μεροληψία και ανακρίβεια:** Οι αγγελίες εργασίας ενδέχεται να αντικατοπτρίζουν προκαταλήψεις όσον αφορά τη γλώσσα, τις απαιτήσεις εργασίας ή τα προτιμώμενα προσόντα. Επιπλέον, ανακρίβειες ή αποκλίσεις στις περιγραφές θέσεων εργασίας μπορεί να οδηγήσουν σε παραπλανητικά συμπεράσματα κατά την ανάλυση.
- **Μεταβλητότητα στη μορφή:** Οι αναρτήσεις θέσεων εργασίας ενδέχεται να ακολουθούν διαφορετικές μορφές σε ιστότοπους και πλατφόρμες, γεγονός που καθιστά δύσκολη την ανάπτυξη μιας ενοποιημένης προσέγγισης για την εξαγωγή και ανάλυση δεδομένων. Οι παραλλαγές στη γλώσσα, την ορολογία και τη διάταξη περιπλέκουν περαιτέρω τη διαδικασία.

- Όγκος και επεκτασιμότητα: Ο τεράστιος όγκος των διαθέσιμων διαδικτυακών αγγελιών εργασίας μπορεί να κατακλύσει τις παραδοσιακές μεθόδους χειροκίνητης ανάλυσης. Απαιτούνται επεκτάσιμες αυτοματοποιημένες λύσεις για την αποτελεσματική επεξεργασία μεγάλων συνόλων δεδομένων.
- Πολύγλωσσα δεδομένα: Στις παγκόσμιες αγορές εργασίας, οι αποσπάσεις μπορεί να είναι διαθέσιμες σε πολλές γλώσσες, προσθέτοντας ένα άλλο επίπεδο πολυπλοκότητας στην ανάλυση δεδομένων. Η μετάφραση γλωσσών και οι τεχνικές ανάκτησης διαγλωσσικών πληροφοριών μπορεί να είναι απαραίτητες για ολοκληρωμένη ανάλυση.
- Απόρρητο και συμμόρφωση: Οι αγγελίες εργασίας ενδέχεται να περιέχουν ευαίσθητες πληροφορίες σχετικά με εταιρείες, άτομα ή αιτούντες εργασία. Η διασφάλιση της συμμόρφωσης με τους κανονισμούς προστασίας δεδομένων και η διατήρηση του απορρήτου είναι πρωταρχικής σημασίας κατά την επεξεργασία και ανάλυση δεδομένων.
- Ενσωμάτωση με άλλες πηγές δεδομένων: Η ενσωμάτωση πληροφοριών από διαδικτυακές αναρτήσεις θέσεων εργασίας με άλλα σχετικά σύνολα δεδομένων, όπως στατιστικά στοιχεία αγοράς εργασίας ή οικονομικά δεδομένα εταιρειών, μπορεί να προσφέρει πλουσιότερες πληροφορίες. Ωστόσο, η διασφάλιση της συμβατότητας και της συνέπειας των δεδομένων μεταξύ των πηγών μπορεί να είναι πρόκληση.

Η αντιμετώπιση αυτών των προκλήσεων απαιτεί έναν συνδυασμό προηγμένων τεχνικών επεξεργασίας δεδομένων, τεχνογνωσίας στον τομέα και ισχυρής υπολογιστικής υποδομής. Η συνεργασία μεταξύ επιστημόνων δεδομένων, εμπειρογνομόνων τομέα και ενδιαφερομένων είναι απαραίτητη για την ανάπτυξη αποτελεσματικών λύσεων, την ανάλυση και την επεξεργασία πληροφοριών από τις διαδικτυακές αναρτήσεις θέσεων εργασίας με ακρίβεια και αποτελεσματικότητα.

ΚΕΦΑΛΑΙΟ 3: Η ΑΝΑΠΤΥΞΗ ΣΥΣΤΗΜΑΤΩΝ ΓΙΑ ΤΗΝ ΑΝΑΛΥΣΗ ΑΓΓΕΛΙΩΝ ΕΡΓΑΣΙΑΣ

3.1. Χρήση μοντέλων Machine Learning και της τεχνίτης νοημοσύνης για την κατηγοριοποίηση και ανάλυση των διαδικτυακών αγγελιών εργασίας

Η εποχή του Διαδικτύου συνεχίζει να αναπτύσσεται, να παράγει έναν ασύλληπτο όγκο δεδομένων κάθε δευτερόλεπτο, ο όγκος και η πολυπλοκότητα των διαθέσιμων πληροφοριών θέτουν σημαντικές προκλήσεις για αποτελεσματική ανάλυση και οργάνωση. Η μη αυτόματη κατηγοριοποίηση και ανάλυση μεγάλων συνόλων δεδομένων δεν είναι μόνο χρονοβόρα αλλά και επιρρεπής σε ανθρώπινο λάθος. Ευτυχώς, οι εξελίξεις στη μηχανική μάθηση έχουν φέρει επανάσταση στον τρόπο με τον οποίο χειριζόμαστε δεδομένα αυτοματοποιώντας εργασίες.

Τα μοντέλα μηχανικής μάθησης, ιδιαίτερα εκείνα που βασίζονται σε αλγόριθμους όπως η βαθιά μάθηση και η επεξεργασία φυσικής γλώσσας, προσφέρουν ισχυρά εργαλεία για αυτόματη κατηγοριοποίηση και ανάλυση δεδομένων. Εκπαιδεύοντας αυτά τα μοντέλα σε επισημασμένα σύνολα δεδομένων, μπορούν να μάθουν να αναγνωρίζουν μοτίβα, να εξάγουν σχετικά χαρακτηριστικά και να λαμβάνουν τεκμηριωμένες αποφάσεις χωρίς ρητές οδηγίες προγραμματισμού.

Οι εφαρμογές της κατηγοριοποίησης και ανάλυσης δεδομένων βάσει της μηχανικής μάθησης αποφέρουν απτά οφέλη σε διάφορους τομείς. Από την τμηματοποίηση πελατών στο *marketing* έως τον εντοπισμό απάτης στα χρηματοοικονομικά. Οι λύσεις που υποστηρίζονται από *ML* ενισχύουν τις διαδικασίες λήψης αποφάσεων, βελτιώνουν την αποτελεσματικότητα και ξεκλειδώνουν νέες ιδέες που οδηγούν στην καινοτομία και το ανταγωνιστικό πλεονέκτημα.

Συνοπτικά, αυτή η εισαγωγή θέτει τις βάσεις για μια ολοκληρωμένη εξερεύνηση της χρήσης μοντέλων μηχανικής εκμάθησης για την αυτόματη κατηγοριοποίηση και ανάλυση δεδομένων. Υιοθετώντας, αυτές τις τεχνολογίες και μεθοδολογίες, οι οργανισμοί μπορούν να αξιοποιήσουν τα στοιχεία δεδομένων τους πιο αποτελεσματικά, δίνοντάς τους τη δυνατότητα να λαμβάνουν τεκμηριωμένες αποφάσεις και να παραμένουν μπροστά σε έναν κόσμο που επικεντρώνεται όλο και περισσότερο στα δεδομένα.

3.1.1. Ορισμός της Μηχανικής Μάθησης

Η μηχανική μάθηση είναι ένας τομέας της επιστήμης των υπολογιστών που επικεντρώνεται στο να δίνει τη δυνατότητα στους υπολογιστές να μαθαίνουν και να βελτιώνονται χωρίς ρητό προγραμματισμό. Αναλύοντας μεγάλες ποσότητες δεδομένων, γνωστά ως δεδομένα εκπαίδευσης, οι αλγόριθμοι μηχανικής μάθησης μπορούν να προσδιορίσουν μοτίβα και σχέσεις. Αυτά τα μοτίβα χρησιμοποιούνται στη συνέχεια για να γίνουν προβλέψεις ή βοηθούν στην λήψη αποφάσεων.

Στις μέρες μας έχουμε παρατηρήσει μια γρήγορη ανάπτυξη λόγω της αυξανόμενης διαθεσιμότητας δεδομένων και της υπολογιστικής ισχύος. Η μηχανική μάθηση (*ML*) έχει γίνει μια μεταμορφωτική δύναμη σε όλους τους κλάδους, φέρνοντας επανάσταση στον τρόπο με τον οποίο οι υπολογιστές μαθαίνουν και αλληλοεπιδρούν με τον κόσμο. Ως υποπεδίο της τεχνητής νοημοσύνης (*AI*), η *ML* εξουσιοδοτεί τις μηχανές να εξάγουν γνώση από δεδομένα, να αναγνωρίζουν μοτίβα και να κάνουν προβλέψεις χωρίς ρητό προγραμματισμό. Αυτή η δυνατότητα ξεκλειδώνει ένα τεράστιο δυναμικό για εφαρμογές που κυμαίνονται από την αναγνώριση προσώπου σε smartphone έως το φιλτράρισμα ανεπιθύμητων μηνυμάτων μέσω *email*. (Géron, 2017)

Η μηχανική μάθηση (*ML*) είναι ένας κλάδος της τεχνητής νοημοσύνης (*AI*) και της επιστήμης των υπολογιστών που εστιάζει στη χρήση δεδομένων και αλγορίθμων για να επιτρέψει στην τεχνητή νοημοσύνη να μιμείται τον τρόπο που μαθαίνουν οι άνθρωποι, βελτιώνοντας σταδιακά την ακρίβειά της. (IBM, n.d.) Η μηχανική μάθηση περιστρέφεται γύρω από την έννοια των αλγορίθμων. Αυτοί οι αλγόριθμοι είναι ουσιαστικά σύνολα οδηγιών που επεξεργάζονται δεδομένα και μαθαίνουν από αυτά. Η διαδικασία μάθησης συνήθως περιλαμβάνει την τροφοδοσία του αλγόριθμου με ένα μεγάλο σύνολο δεδομένων, γνωστό ως δεδομένα εκπαίδευσης. (Brownlee, 2024)

Όπως αναφέραμε και παραπάνω στον ορισμό η μηχανική μάθηση (*ML*) είναι ένας τύπος τεχνητής νοημοσύνης που επικεντρώνεται στην κατασκευή συστημάτων υπολογιστών που μαθαίνουν από δεδομένα. Το ευρύ φάσμα τεχνικών που περιλαμβάνει η *ML* επιτρέπει στις εφαρμογές λογισμικού να βελτιώνουν την απόδοσή τους με την πάροδο του χρόνου. Οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται να βρίσκουν σχέσεις και μοτίβα στα δεδομένα. Χρησιμοποιούν ιστορικά δεδομένα ως είσοδο για να κάνουν προβλέψεις, για να ταξινομήσουν πληροφορίες, να συγκεντρώνουν σημεία δεδομένων και να μειώνουν τις διαστάσεις. Ακόμη, βοηθούν στη δημιουργία νέου περιεχομένου, όπως αποδεικνύεται από νέες εφαρμογές που

τροφοδοτούνται με *ML*, όπως το *ChatGPT*, το *Dall-E 2* και το *GitHub Copilot*. Η μηχανική μάθηση είναι ευρέως εφαρμόσιμη σε πολλούς κλάδους. Οι μηχανές συστάσεων, για παράδειγμα, χρησιμοποιούνται από το ηλεκτρονικό εμπόριο, τα μέσα κοινωνικής δικτύωσης και τους ειδησεογραφικούς οργανισμούς για να προτείνουν περιεχόμενο με βάση την προηγούμενη συμπεριφορά ενός πελάτη. Οι αλγόριθμοι μηχανικής μάθησης περιλαμβάνουν τον εντοπισμό απάτης, το φιλτράρισμα ανεπιθύμητων μηνυμάτων, τον εντοπισμό απειλών κακόβουλου λογισμικού, την προγνωστική συντήρηση και την αυτοματοποίηση επιχειρηματικών διαδικασιών.

Η μηχανική μάθηση μπορεί να αποτελεί ένα ισχυρό εργαλείο για την επίλυση προβλημάτων, τη βελτίωση των επιχειρηματικών λειτουργιών και την αυτοματοποίηση εργασιών. Όμως, αποτελεί μια πολύπλοκη τεχνολογία, που απαιτεί βαθιά τεχνογνωσία και σημαντικούς πόρους. Η επιλογή του σωστού αλγόριθμου για μια εργασία απαιτεί ισχυρή κατανόηση των μαθηματικών και της στατιστικής. Η εκπαίδευση αλγορίθμων μηχανικής εκμάθησης συχνά περιλαμβάνει μεγάλες ποσότητες δεδομένων καλής ποιότητας για την παραγωγή ακριβών αποτελεσμάτων. Τα ίδια τα αποτελέσματα μπορεί να είναι δύσκολο να κατανοηθούν ιδιαίτερα τα αποτελέσματα που παράγονται από πολύπλοκους αλγόριθμους, όπως τα νευρωνικά δίκτυα βαθιάς μάθησης που έχουν σχεδιαστεί σύμφωνα με τον ανθρώπινο εγκέφαλο. Και τα μοντέλα *ML* μπορεί να είναι δαπανηρά για την εκτέλεση και τον συντονισμό. Οι περισσότεροι οργανισμοί είτε άμεσα είτε έμμεσα χρησιμοποιούν την τεχνητή νοημοσύνη και την μηχανική μάθηση. Καθώς, αποτελούν μέρος των στρατηγικών πληροφορικής για πορεία των επιχειρήσεων τους. (Burns, 2024) Η μηχανική μάθηση έχει γίνει αναπόσπαστο μέρος του επιχειρηματικού λογισμικού που διαχειρίζεται τους οργανισμούς. Ακολουθούν ορισμένα παραδείγματα του τρόπου με τον οποίο διάφοροι κλάδοι χρησιμοποιούν την *ML*:

- 📌 Επιχειρηματική ευφυΐα: Το *BI* και το λογισμικό πρόβλεψης ανάλυσης χρησιμοποιούν αλγόριθμους μηχανικής μάθησης, συμπεριλαμβανομένης της γραμμικής παλινδρόμησης και της λογιστικής παλινδρόμησης, για τον εντοπισμό σημαντικών σημείων δεδομένων, μοτίβων και ανωμαλιών σε μεγάλα σύνολα δεδομένων.
- 📌 Διαχείριση πελατειακών σχέσεων: Οι βασικές εφαρμογές της μηχανικής εκμάθησης στο *CRM* περιλαμβάνουν την ανάλυση δεδομένων πελατών για την τμηματοποίηση πελατών, την πρόβλεψη συμπεριφορών όπως η απόρριψη, η υποβολή προτάσεων, η

προσαρμογή των τιμών, η βελτιστοποίηση καμπανιών *email*, η παροχή υποστήριξης *chatbots* και ο εντοπισμός απάτης.

- ✚ Ασφάλεια και συμμόρφωση: Προηγμένοι αλγόριθμοι όπως η ανίχνευση ανωμαλιών και οι τεχνικές μηχανών υποστήριξης διανυσμάτων (*SVM*) προσδιορίζουν την κανονική συμπεριφορά και τις αποκλίσεις, κάτι που είναι ζωτικής σημασίας για τον εντοπισμό πιθανών απειλών στον κυβερνοχώρο. Τα *SVM* βρίσκουν την καλύτερη γραμμή που χωρίζει τα δεδομένα σε διαφορετικές ομάδες που χωρίζονται με όσο το δυνατόν περισσότερο χώρο.
- ✚ Πληροφοριακά συστήματα ανθρώπινου δυναμικού: Τα μοντέλα *ML* απλοποιούν τη διαδικασία πρόσληψης φιλτράροντας τις εφαρμογές και εντοπίζοντας τους καλύτερους υποψηφίους για μια ανοιχτή θέση.
- ✚ Διαχείριση εφοδιαστικής αλυσίδας: Οι τεχνικές μηχανικής εκμάθησης βελτιστοποιούν τα επίπεδα αποθέματος, βελτιώνουν την επιλογή προμηθευτών και αντιμετωπίζουν προληπτικά τις διακοπές της εφοδιαστικής αλυσίδας.
- ✚ Επεξεργασία φυσικής γλώσσας : Τα μοντέλα *ML* επιτρέπουν σε εικονικούς βοηθούς όπως η *Alexa*, το *Google Assistant* και το *Siri* να ερμηνεύουν και να ανταποκρίνονται στην ανθρώπινη γλώσσα. (Linda Tucci, 2024)

3.1.2. Η λειτουργία της τεχνολογίας της Μηχανικής Μάθησης

Στα μέσα του 20ου αιώνα η μηχανική μάθηση κατέχει τον κεντρικό ρόλο στην ανθρώπινη κοινωνία, έχοντας ως πρωτοπόρους της τεχνητής νοημοσύνης τον *Walter Pitts*, τον *Warren McCulloch*, τον *Alan Turing* και τον *John von Neumann*. Καθώς, έθεσαν τις βάσεις για την εκπαίδευση των μηχανών ώστε να μαθαίνουν από τα δεδομένα και να βελτιώνονται με την πάροδο του χρόνου. Έχει επιτραπεί στους οργανισμούς να αυτοματοποιούν εργασίες

ρουτίνας που εκτελούνταν προηγουμένως από ανθρώπους. Επίσης, μπορεί να εκτελέσει χειροκίνητες εργασίες που είναι πέρα από την ικανότητά του ανθρώπου. Για παράδειγμα, η επεξεργασία των τεράστιων ποσοτήτων δεδομένων που παράγονται σήμερα από ψηφιακές συσκευές. Η ικανότητα της μηχανικής μάθησης να εξάγει μοτίβα και γνώσεις από τεράστια σύνολα δεδομένων έχει γίνει ανταγωνιστική διαφοροποίηση σε τομείς που κυμαίνονται από τη χρηματοδότηση, το λιανικό εμπόριο έως την υγειονομική περίθαλψη και την επιστημονική ανακάλυψη. Πολλές από τις κορυφαίες εταιρείες του σήμερα, συμπεριλαμβανομένου του *Facebook*, της *Google* και της *Uber*, καθιστούν τη μηχανική εκμάθηση κεντρικό μέρος των εργασιών τους.

Καθώς, ο όγκος των δεδομένων που παράγονται από τις σύγχρονες κοινωνίες συνεχίζει να πολλαπλασιάζεται, η μηχανική μάθηση πιθανότατα θα γίνει ακόμη πιο ζωτική για τους ανθρώπους και απαραίτητη για την ίδια τη μηχανική νοημοσύνη. Η τεχνολογία όχι μόνο μας βοηθά να κατανοήσουμε τα δεδομένα που δημιουργούμε, αλλά ενισχύει περαιτέρω τις δυνατότητες εκμάθησης της *ML* που βασίζονται σε δεδομένα.

Η μηχανική μάθηση βασίζεται στην ιδέα ότι οι υπολογιστές μπορούν να "μάθουν" από δεδομένα, χωρίς να χρειάζεται να προγραμματιστούν ρητά για κάθε εργασία. Η διαδικασία λαμβάνει χώρα σε 5 βασικά στάδια:

- I. Συλλογή δεδομένων: Το πρώτο βήμα είναι η συγκέντρωση ενός μεγάλου όγκου δεδομένων, σχετικών με την εργασία που θέλουμε να εκτελέσει το σύστημα μηχανικής μάθησης. Τα δεδομένα αυτά μπορούν να είναι δομημένα (π.χ. πίνακες) ή μη δομημένα (π.χ. κείμενο, εικόνες).
- II. Προ επεξεργασία δεδομένων: Τα δεδομένα συχνά χρειάζονται επεξεργασία πριν τροφοδοτήσουν το σύστημα. Αυτό μπορεί να περιλαμβάνει καθαρισμό (αφαίρεση σφαλμάτων ή ατελών δεδομένων), τυποποίηση (μετατροπή δεδομένων σε συμβατή μορφή) και ομαλοποίηση (εξισορρόπηση της κλίμακας διαφορετικών χαρακτηριστικών).
- III. Εκπαίδευση μοντέλου: Στη συνέχεια, επιλέγεται ένας αλγόριθμος μηχανικής μάθησης και τροφοδοτείται με τα προ επεξεργασμένα δεδομένα. Ο αλγόριθμος "μαθαίνει" από τα δεδομένα, εντοπίζοντας πρότυπα και σχέσεις, και δημιουργεί ένα "μοντέλο".

- IV. Αξιολόγηση και χρήση μοντέλου: Το μοντέλο αξιολογείται με ένα σύνολο "δοκιμαστικών" δεδομένων, για να ελεγχθεί η ακρίβειά του. Εάν η απόδοση είναι ικανοποιητική, το μοντέλο μπορεί να χρησιμοποιηθεί για να κάνει προβλέψεις ή να λαμβάνει αποφάσεις σε νέα δεδομένα.
- V. Βελτιστοποίηση του αλγορίθμου: Αν η απόδοση του αλγορίθμου δεν είναι ικανοποιητική, μπορεί να τροποποιηθεί ή να εκπαιδευτεί ξανά με διαφορετικά δεδομένα.

Η κλασική μηχανική μάθηση συχνά κατηγοριοποιείται με βάση το πώς ένας αλγόριθμος μαθαίνει να γίνεται πιο ακριβής στις προβλέψεις του. Υπάρχουν τέσσερις βασικοί τύποι μηχανικής μάθησης :

✚ Η εποπτευόμενη μάθηση: Οι επιστήμονες δεδομένων παρέχουν αλγόριθμους με επισημασμένα δεδομένα εκπαίδευσης και ορίζουν τις μεταβλητές που θέλουν να αξιολογήσει ο αλγόριθμος. Τόσο η είσοδος όσο και η έξοδος του αλγορίθμου καθορίζονται στην εποπτευόμενη μάθηση. Αρχικά, οι περισσότεροι αλγόριθμοι μηχανικής μάθησης λειτουργούσαν με εποπτευόμενη μάθηση, αλλά οι μη εποπτευόμενες προσεγγίσεις γίνονται δημοφιλείς. Σε διάφορες εργασίες οι αλγόριθμοί εποπτευόμενης μάθησης χρησιμοποιούν την δυαδική ταξινόμηση όπου χωρίζει τα δεδομένα σε δυο κατηγορίες, την πολυταξική ταξινόμηση όπου επιλέγει ανάμεσα σε περισσότερους από δυο τύπους απαντήσεων. Καθώς, συνδυάζει προβλέψεις πολλαπλών μοντέλων *ML* για την παραγωγή μιας πιο ακριβής πρόβλεψης. Τέλος, οι αλγόριθμοι εποπτευόμενης μάθησης χρησιμοποιούν την μοντελοποίηση παλινδρόμησης όπου προβλέπει συνεχείς τιμές σύμφωνα με τα δεδομένα.

✚ Η μάθηση χωρίς επίβλεψη: Οι αλγόριθμοι μηχανικής εκμάθησης χωρίς επίβλεψη δεν απαιτούν επισημάνση δεδομένων. Αφαιρούν τα δεδομένα χωρίς ετικέτα για να αναζητήσουν μοτίβα που μπορούν να χρησιμοποιηθούν για την ομαδοποίηση σημείων δεδομένων σε υποσύνολα. Οι περισσότεροι τύποι βαθιάς μάθησης, συμπεριλαμβανομένων των νευρωνικών δικτύων, είναι αλγόριθμοι χωρίς επίβλεψη. Οι αλγόριθμοι μάθησης χωρίς επίβλεψη είναι καλοί για εργασίες όπως είναι η ομαδοποίηση για των διαχωρισμός του συνόλου δεδομένων σε ομάδες με βάση την

ομοιότητα χρησιμοποιώντας αλγόριθμους ομαδοποίησης. Επιπλέον, βοηθάει στην ανίχνευση ανωμαλιών για τον προσδιορισμό ασυνήθιστων σημείων δεδομένων σε ένα σύνολο δεδομένων χρησιμοποιώντας αλγόριθμους ανίχνευσης ανωμαλιών. Καθώς, συμβάλουν στην ανακάλυψη συνόλων στοιχείων σε ένα σύνολο δεδομένων που εμφανίζονται συχνά μαζί χρησιμοποιώντας εξόρυξη κανόνων συσχέτισης και στην μείωση διαστάσεων όπως είναι η μείωση του αριθμού των μεταβλητών σε ένα σύνολο δεδομένων χρησιμοποιώντας τεχνικές μείωσης διαστάσεων.

✚ Η ημιεποπτευόμενη μάθηση : Η ημιεποπτευόμενη μάθηση λειτουργεί τροφοδοτώντας μια μικρή ποσότητα δεδομένων εκπαίδευσης με ετικέτα σε έναν αλγόριθμο. Από αυτά τα δεδομένα, ο αλγόριθμος μαθαίνει τις διαστάσεις του συνόλου δεδομένων, τις οποίες μπορεί στη συνέχεια να εφαρμόσει σε νέα δεδομένα χωρίς ετικέτα. Η απόδοση των αλγορίθμων συνήθως βελτιώνεται όταν εκπαιδεύονται σε σύνολα δεδομένων με ετικέτα. Αλλά η επισήμανση δεδομένων μπορεί να είναι χρονοβόρα και δαπανηρή. Αυτός ο τύπος μηχανικής μάθησης επιτυγχάνει μια ισορροπία μεταξύ της ανώτερης απόδοσης της εποπτευόμενης μάθησης και της αποτελεσματικότητας της μάθησης χωρίς επίβλεψη. Αυτό το είδος μάθησης μπορεί να χρησιμοποιηθεί στους ακόλουθους τομείς όπως είναι η μηχανική μετάφραση όπου διδάσκει αλγόριθμους για τη μετάφραση γλώσσας με βάση λιγότερο από ένα πλήρες λεξικό λέξεων, στην ανίχνευση απάτης και σε δεδομένα επισήμανσης. Όπου οι αλγόριθμοι εκπαιδεύονται σε μικρά σύνολα δεδομένων μαθαίνουν να εφαρμόζουν αυτόματα ετικέτες δεδομένων σε μεγαλύτερα σύνολα.

✚ Η ενισχυτική μάθηση : Λειτουργεί προγραμματίζοντας έναν αλγόριθμο για έναν ξεχωριστό στόχο και ένα προδιαγεγραμμένο σύνολο κανόνων για την επίτευξη αυτού του στόχου. Ένας επιστήμονας δεδομένων θα προγραμματίσει επίσης τον αλγόριθμο ώστε να αναζητά θετικές ανταμοιβές για την εκτέλεση μιας ενέργειας που είναι ωφέλιμη για την επίτευξη του απώτερου στόχου της και να αποφύγει ενέργειας που μπορεί να τον απομακρύνει περισσότερο από τον στόχο. Η ενισχυτική μάθηση χρησιμοποιείται συχνά σε τομείς όπως είναι η ρομποτική μαθαίνοντας τα ρομπότ να εκτελούν εργασίες στον φυσικό κόσμο, σε βιντεοπαιχνίδι και στην διαχείριση πόρων ωφελώντας τις επιχειρήσεις να σχεδιάσουν την κατανομή των πόρων.

Αξίζει να σημειωθεί ότι, ο τύπος αλγορίθμου που επιλέγουν οι επιστήμονες δεδομένων εξαρτάται από τη φύση των δεδομένων. Πολλοί από τους αλγόριθμους και τις τεχνικές δεν περιορίζονται μόνο σε έναν από τους κύριους τύπους *ML*. Συχνά προσαρμόζονται σε πολλούς τύπους, ανάλογα με το πρόβλημα που πρέπει να επιλυθεί και το σύνολο δεδομένων. Για παράδειγμα, αλγόριθμοι βαθιάς μάθησης, όπως τα συνελκτικά νευρωνικά δίκτυα και τα επαναλαμβανόμενα νευρωνικά δίκτυα, χρησιμοποιούνται σε εποπτευόμενες, μη εποπτευόμενες και ενισχυτικές εργασίες μάθησης, με βάση το συγκεκριμένο πρόβλημα και τη διαθεσιμότητα των δεδομένων. (Linda Tucci, 2024)

3.1.3. Τα προνόμια που προσφέρει η Μηχανική Μάθηση στις μέρες μας

Η μηχανική μάθηση (*Machine Learning*) αποτελεί έναν κλάδο της επιστήμης υπολογιστών που εστιάζει στην ανάπτυξη αλγορίθμων ικανών να "μαθαίνουν" από δεδομένα. Σε αντίθεση με παραδοσιακές προσεγγίσεις προγραμματισμού, όπου οι αλγόριθμοι υλοποιούνται με σαφείς οδηγίες, η μηχανική μάθηση επιτρέπει στους υπολογιστές να βελτιστοποιούν την απόδοσή τους αυτόνομα, βασιζόμενοι σε εμπειρικά δεδομένα. Την σήμερα ημέρα, η μηχανική μάθηση έχει φέρει επανάσταση σε διάφορους κλάδους καθώς συνεχίζει να είναι η κινητήρια δύναμη πίσω από τις καινοτόμες εξελίξεις. Η ικανότητά του να αναλύει τεράστια δεδομένα και να κάνει προβλέψεις χωρίς ρητό προγραμματισμό έχει οδηγήσει σε πολλά πλεονεκτήματα τα οποία είναι :

- ✚ Αυτοματοποίηση: Ένα από τα σημαντικά πλεονεκτήματα της μηχανικής μάθησης είναι οι προηγμένες δυνατότητες αυτοματισμού της. Οι αλγόριθμοι μηχανικής μάθησης έχουν γίνει πιο εξελιγμένοι, επιτρέποντας την αυτοματοποίηση πολύπλοκων εργασιών και διαδικασιών σε διάφορους κλάδους. Αυτός ο αυτοματισμός μειώνει την ανάγκη για ανθρώπινη παρέμβαση σε επαναλαμβανόμενες και χρονοβόρες δραστηριότητες, οδηγώντας σε αυξημένη παραγωγικότητα και αποδοτικότητα κόστους. Στην κατασκευή, η μηχανική εκμάθηση διευκολύνει την αυτοματοποίηση των γραμμών συναρμολόγησης και τον ποιοτικό έλεγχο, οδηγώντας σε εξορθολογισμένες διαδικασίες παραγωγής. Στην εξυπηρέτηση πελατών, τα *chatbots* που υποστηρίζονται από μηχανική εκμάθηση χειρίζονται τα ερωτήματα και την υποστήριξη πελατών, παρέχοντας άμεσες απαντήσεις και βελτιώνοντας τις εμπειρίες των χρηστών. Στα χρηματοοικονομικά, η μηχανική εκμάθηση αυτοματοποιημένης αξιολόγησης

κινδύνου, ανίχνευση απάτης και ανάληψη πιστώσεων, ενισχύοντας την ταχύτητα και την ακρίβεια των χρηματοοικονομικών υπηρεσιών.

- ✚ Εύρος Εφαρμογών: Ένα άλλο πλεονέκτημα της μηχανικής μάθησης έγκειται στο συνεχώς διευρυμένο φάσμα εφαρμογών της σε διάφορους κλάδους. Καθώς, η τεχνολογία εξελίσσεται, γίνεται όλο και πιο ευέλικτη και προσαρμόσιμη για την αντιμετώπιση σύνθετων προκλήσεων και ευκαιριών. Η μηχανική μάθηση βρίσκει εφαρμογές στην υγειονομική περίθαλψη, τα οικονομικά, το μάρκετινγκ, την κατασκευή, τις μεταφορές κ.λπ. Για παράδειγμα, στα χρηματοοικονομικά, εξουσιοδοτεί την αξιολόγηση κινδύνου, τον εντοπισμό απάτης και τις αλγοριθμικές συναλλαγές.
- ✚ Προσδιορίζει τάσεις και μοτίβα: Η μηχανική μάθηση μπορεί να εντοπίσει τάσης και πρότυπα σε διάφορους κλάδους. Με την εκθετική αύξηση των δεδομένων, οι αλγόριθμοι μηχανικής μάθησης έχουν γίνει ακόμη πιο ισχυροί στην ανάλυση τεράστιων και πολύπλοκων συνόλων δεδομένων. Με την επεξεργασία αυτών των δεδομένων, οι αλγόριθμοι *ML* μπορούν να αποκαλύψουν πολύτιμες πληροφορίες, συσχετίσεις και τάσεις που μπορεί να μην είναι εμφανείς μέσω των παραδοσιακών μεθόδων. Οι επιχειρήσεις επωφελούνται πάρα πολύ από αυτή τη δυνατότητα, καθώς η μηχανική μάθηση τους δίνει τη δυνατότητα να λαμβάνουν αποφάσεις βάσει δεδομένων, να βελτιστοποιούν τις λειτουργίες τους και να αποκτούν ανταγωνιστικό πλεονέκτημα.
- ✚ Αποτελεσματικός χειρισμός δεδομένων: Με την εκθετική αύξηση των δεδομένων, οι παραδοσιακές μέθοδοι επεξεργασίας δεδομένων έχουν γίνει ανεπαρκείς. Οι αλγόριθμοι μηχανικής μάθησης υπερέχουν στην επεξεργασία, την ανάλυση και την εξαγωγή πολύτιμων πληροφοριών από μεγάλους όγκους δεδομένων γρήγορα και με ακρίβεια. Η μηχανική μάθηση διαχειρίζεται αποτελεσματικά χρηματοοικονομικά αρχεία με οικονομικά δεδομένα για την εκτίμηση κινδύνου και ανίχνευσης απάτης.
- ✚ Καινοτομία: Καθώς, η τεχνολογία προχωρά, οι αλγόριθμοι μηχανικής μάθησης γίνονται πιο ισχυροί, εξελιγμένοι και ικανοί να αντιμετωπίζουν όλο και πιο περίπλοκες

προκλήσεις. Οι ερευνητές και οι προγραμματιστές συνεχώς καινοτομούν και βελτιώνουν αυτούς τους αλγόριθμους, επεκτείνοντας τις δυνατότητές τους και τις πιθανές εφαρμογές τους. Το εύρος της βελτίωσης στη μηχανική μάθηση επωφελή διάφορους κλάδους με αποτέλεσμα η μηχανική εκμάθηση να ενισχύει τις ικανότητες πλοήγησης, ασφάλειας και λήψης αποφάσεων.

✚ Βελτιωμένη εμπειρία στις ηλεκτρονικές αγορές και ποιοτική εκπαίδευση: Η μηχανική εκμάθηση παρουσιάζει σημαντικό πλεονέκτημα ενισχύοντας την εμπειρία αγορών μέσω διαδικτύου και την ποιοτική εκπαίδευση. Στο ηλεκτρονικό εμπόριο, οι αλγόριθμοι μηχανικής μάθησης τροφοδοτούν εξελιγμένα συστήματα προτάσεων που αναλύουν τη συμπεριφορά των χρηστών, το ιστορικό αγορών και τις προτιμήσεις για να προσφέρουν εξατομικευμένες προτάσεις προϊόντων.

✚ Χειρισμός Πολυδιάστατων και Πολυποίκιλων Δεδομένων: Η μηχανική μάθηση προσφέρει μία εξαιρετική ικανότητα στον χειρισμό πολυδιάστατα και πολυποικιλιακά δεδομένα. Με τη συνεχή ανάπτυξη των πηγών και των τύπων δεδομένων, οι παραδοσιακές μέθοδοι επεξεργασίας δεδομένων αγωνίζονται να αναλύσουν και να εξάγουν πληροφορίες από πολύπλοκα σύνολα δεδομένων αποτελεσματικά.

Οι αλγόριθμοι μηχανικής μάθησης υπερέχουν στην επεξεργασία διαφορετικών μορφών δεδομένων, όπως κείμενο, εικόνες, ήχος και δεδομένα αισθητήρων, ενώ λαμβάνονται υπόψη πολλαπλές διαστάσεις και συσχετίσεις. Στον χρηματοοικονομικό κλάδο, μπορεί να αναλύσει δεδομένα από διάφορες πηγές, όπως οι τάσεις της αγοράς, οι οικονομικοί δείκτες και τα μέσα κοινωνικής δικτύωσης, για να λάβει τεκμηριωμένες επενδυτικές αποφάσεις. (Vadapalli, 2023) Αξίζει να αναφέρουμε ότι η μηχανική μάθηση αποτελεί ένα ισχυρό εργαλείο με πολλαπλές δυνατότητες. Η υιοθέτησή της μπορεί να προσφέρει σημαντικά οφέλη σε επιχειρήσεις και οργανισμούς, οδηγώντας σε αυξημένη παραγωγικότητα, βελτιωμένη λήψη αποφάσεων και ανταγωνιστικό πλεονέκτημα.

3.1.4. Η χρήση της μηχανικής μάθησης στην εξαγωγή πληροφορίες από διαδικτυακές αγγελίες εργασίας

Η μηχανική μάθηση μπορεί να εξάγει πολλές πληροφορίες από διαδικτυακές αγγελίες εργασίας χρησιμοποιώντας διάφορες τεχνικές και αλγόριθμους. Ας δούμε μερικούς τρόπους με τους οποίους η μηχανική μάθηση μπορεί να επεξεργαστεί αυτές τις αγγελίες:

- Κατηγοριοποίηση Θέσεων Εργασιών: Ένας αλγόριθμος μηχανικής μάθησης μπορεί να εκπαιδευτεί να αναγνωρίζει τις διάφορες κατηγορίες εργασιών μέσα στις αγγελίες, όπως π.χ. προγραμματιστής, σχεδιαστής, διαχειριστής κλπ.
- Εξαγωγή Κλειδιών Λέξεων: Με τη χρήση τεχνικών επεξεργασίας φυσικής γλώσσας, η μηχανική μάθηση μπορεί να εξάγει τις κύριες λέξεις-κλειδιά από τις αγγελίες εργασίας, που μπορούν να παρέχουν ενδείξεις σχετικά με τον τύπο της εργασίας και τις απαιτούμενες δεξιότητες.
- Πρόβλεψη Αποδοχής Εργασίας: Με τη χρήση δεδομένων από προηγούμενες αγγελίες και τις συναφείς αποδοχές εργασίας, ένα μοντέλο μηχανικής μάθησης μπορεί να εκπαιδευτεί να προβλέπει την πιθανότητα αποδοχής μιας συγκεκριμένης εργασίας από έναν υποψήφιο βάσει των χαρακτηριστικών της αγγελίας και του βιογραφικού του.
- Προβλέψεις Εξέλιξης Αγοράς Εργασίας : Μέσω της συλλογής και ανάλυσης αγγελιών εργασίας από διάφορες εταιρείες, η μηχανική μάθηση μπορεί να αναδειξεί τις τάσεις της αγοράς εργασίας, τις απαιτούμενες δεξιότητες και τις αμοιβές, παρέχοντας σημαντικές πληροφορίες για ανταγωνιστικές αναλύσεις.

Αυτοί είναι μερικοί τρόποι με τους οποίους η μηχανική μάθηση μπορεί να εξάγει πληροφορίες από διαδικτυακές αγγελίες εργασίας. Είναι σημαντικό να σημειωθεί, ότι η αποτελεσματικότητα αυτών των μεθόδων εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των μοντέλων.

3.2.Ανάλυση αγγελιών εργασίας με την χρήση της τεχνητής νοημοσύνης

Η ανάλυση αγοράς εργασίας με την βοήθεια της τεχνητής νοημοσύνης έχει παρουσιάσει μια σημαντική μεταμόρφωση, με την τεχνολογία να διαδραματίζει ολοένα

και πιο κρίσιμο ρόλο. Η τεχνητή νοημοσύνη (AI) χρησιμοποιείται σε διάφορες πτυχές της διαδικασίας πρόσληψης και η ανάλυση αγγελιών εργασίας. Αυτή η ανάλυση περιλαμβάνει την εξαγωγή και την ερμηνεία πληροφοριών από τις περιγραφές θέσεων εργασίας για την απόκτηση πολύτιμων γνώσεων σχετικά με τις δεξιότητες, τα προσόντα και τις τάσεις που απαιτούνται από την αγορά εργασίας. Με τη μόχλευση της τεχνητής νοημοσύνης, οι επιχειρήσεις και όσοι αναζητούν εργασία μπορούν να ξεκλειδώσουν μια πληθώρα γνώσεων, οδηγώντας σε πιο ενημερωμένες αποφάσεις και βελτιωμένα αποτελέσματα.

Η τεχνητή νοημοσύνη μας δίνει τη δυνατότητα να αναλύουμε τις αγγελίες εργασίας με τρόπο που δεν μπορούν οι παραδοσιακές μέθοδοι. Με την βοήθεια τις :

- **Επεξεργασία Φυσικής Γλώσσας :** Η τεχνητή νοημοσύνη με δυνατότητες *NLP* μπορεί να αναλύει τις περιγραφές θέσεων εργασίας, να προσδιορίζει λέξεις-κλειδιά, δεξιότητες, εκπαιδευτικές απαιτήσεις και επίπεδα εμπειρίας. Εργαλεία όπως το *NLTK* και το *spracy* υπερέχουν σε αυτές τις εργασίες, επιτρέποντας στην τεχνητή νοημοσύνη να κατανοήσει τις βασικές απαιτήσεις και το επιθυμητό προφίλ υποψηφίου για κάθε ανάρτηση εργασίας.
- **Μηχανική Μάθηση (ML):** Οι αλγόριθμοι *ML* μπορούν να εκπαιδευτούν σε τεράστια σύνολα δεδομένων αγγελιών εργασίας. Αυτοί οι αλγόριθμοι μπορούν στη συνέχεια να χρησιμοποιηθούν για την κατηγοριοποίηση των θέσεων εργασίας, τον εντοπισμό των αναδυόμενων τάσεων δεξιοτήτων και ακόμη και την πρόβλεψη των μελλοντικών απαιτήσεων της αγοράς εργασίας. Τεχνικές όπως οι Υποστήριξη Διανυσματικές Μηχανές (*SVM*) και τα Τυχαία Δάση χρησιμοποιούνται συχνά για τέτοιες εργασίες, επιτρέποντας στην τεχνητή νοημοσύνη να αποκαλύψει πολύτιμα μοτίβα και ιδέες από μεγάλους όγκους δεδομένων.
- **Deep Learning (DL):** Τα μοντέλα βαθιάς μάθησης, ιδιαίτερα τα επαναλαμβανόμενα νευρωνικά δίκτυα (*RNN*) με μονάδες μακροπρόθεσμης βραχυπρόθεσμης μνήμης (*LSTM*) είναι ιδιαίτερα ικανά στην κατανόηση την ανθρώπινης γλώσσα. Αυτά τα μοντέλα μπορούν να αναλύσουν το πλαίσιο και τις σχέσεις μεταξύ των λέξεων σε μια περιγραφή εργασίας, παρέχοντας μια πιο λεπτή κατανόηση των απαιτούμενων δεξιοτήτων και εμπειρίας σε σύγκριση με απλούστερες τεχνικές *NLP*.

Η ανάλυση αναρτημένων θέσεων εργασίας με τη βοήθεια της τεχνητής νοημοσύνης μπορεί να μας προσφέρει μια πληθώρα από πλεονεκτήματα. Οι ενδιαφερόμενοι που μπορούν να ωφεληθούν σημαντικά είναι :

- ❖ Εργοδότες: Η τεχνητή νοημοσύνη μπορεί να βοηθήσει τους εργοδότες να εντοπίσουν τις πιο σχετικές δεξαμενές ταλέντων αναλύοντας τις δεξιότητες και τα προσόντα που αναφέρονται στις περιγραφές θέσεων εργασίας σε διαφορετικούς κλάδους και τοποθεσίες. Αυτό μπορεί να οδηγήσει σε πιο στοχευμένες προσπάθειες πρόσληψης και σε μείωση του χρόνου για την πρόσληψη.
- ❖ Άτομα που αναζητούν εργασία: Η τεχνητή νοημοσύνη μπορεί να βοηθήσει όσους αναζητούν εργασία να κατανοήσουν τις δεξιότητες και τις εμπειρίες που έχουν μεγαλύτερη ζήτηση στον τομέα τους. Προσαρμόζοντας τα βιογραφικά τους και τις συνοδευτικές τους επιστολές για να τονίσουν αυτές τις επιθυμητές δεξιότητες, τα άτομα που αναζητούν εργασία μπορούν να αυξήσουν σημαντικά τις πιθανότητές τους για συνέντευξη.
- ❖ Υπεύθυνοι χάραξης πολιτικής: Η ανάλυση των αγγελιών εργασίας που βασίζεται σε τεχνητή νοημοσύνη όπου μπορεί να προσφέρει πολύτιμες πληροφορίες για τις τάσεις της αγοράς εργασίας. Επίσης, επιστρώνεται στους υπεύθυνους χάραξης πολιτικής για την ανάπτυξη στοχευμένων προγραμμάτων ανάπτυξης εργατικού δυναμικού και πρωτοβουλίες που αντιμετωπίζουν τα αναδυόμενα κενά δεξιοτήτων.
- ❖ Ερευνητές: Οι ερευνητές μπορούν να αξιοποιήσουν την τεχνητή νοημοσύνη για μεγάλης κλίμακας ανάλυση των αγγελιών εργασίας. Οδηγώντας σε μια βαθύτερη κατανόηση της μεταβαλλόμενης φύσης της εργασίας, των δημογραφικών στοιχείων του εργατικού δυναμικού και του αντίκτυπου της αυτοματοποίησης σε διαφορετικούς τομείς εργασίας.

Παρά τα πιθανά οφέλη, η ανάλυση αγγελιών εργασίας με τεχνητή νοημοσύνη αντιμετωπίζει ορισμένες προκλήσεις όπως :

- ✚ Ποιότητα δεδομένων και προκατάληψη: Η ακρίβεια των πληροφοριών που εξάγονται από την τεχνητή νοημοσύνη εξαρτάται από την ποιότητα και την ποικιλομορφία των δεδομένων που χρησιμοποιούνται για την εκπαίδευση μοντέλων. Τα μεροληπτικά δεδομένα στις περιγραφές θέσεων εργασίας μπορούν να οδηγήσουν σε μεροληπτικά αποτελέσματα στην ανάλυση τεχνητής νοημοσύνης, με αποτέλεσμα να δημιουργούνται ενδεχομένως οι υπάρχουσες ανισότητες στην αγορά εργασίας.
- ✚ Μεταβλητότητα περιγραφής εργασίας :Οι περιγραφές θέσεων εργασίας συχνά στερούνται τυποποίησης, γεγονός που καθιστά δύσκολο για τα μοντέλα τεχνητής νοημοσύνης να εξάγουν με ακρίβεια συνεπείς πληροφορίες. Οι τυποποιημένες μορφές περιγραφής εργασίας και οντολογίες μπορούν να βοηθήσουν στη βελτίωση της ακρίβειας και της αποτελεσματικότητας της ανάλυσης που βασίζεται σε τεχνητή νοημοσύνη.
- ✚ Επεξηγησιμότητα των αποτελεσμάτων: Η κατανόηση του τρόπου με τον οποίο τα μοντέλα τεχνητής νοημοσύνης καταλήγουν στα συμπεράσματά τους είναι ζωτικής σημασίας για την οικοδόμηση εμπιστοσύνης στην ανάλυση αναρτήσεων εργασίας που βασίζεται στην τεχνητή νοημοσύνη. Η εφαρμογή τεχνικών εξηγήσιμης τεχνητής νοημοσύνης (*XAI*) μπορεί να βοηθήσει τους χρήστες να ερμηνεύσουν τα αποτελέσματα και να εξασφαλίσουν δικαιοσύνη στη διαδικασία πρόσληψης.

Η ενσωμάτωση της τεχνητής νοημοσύνης στην ανάλυση της αγοράς εργασίας υπόσχεται τεράστια υποσχέσεις για το μέλλον της εργασίας. Παρέχοντας πολύτιμες γνώσεις σχετικά με τις δεξιότητες, τα προσόντα και τις τάσεις της αγοράς. Η τεχνητή νοημοσύνη μπορεί να εξουσιοδοτήσει τόσο τους εργοδότες όσο και τα άτομα που αναζητούν εργασία να λαμβάνουν τεκμηριωμένες αποφάσεις στη σημερινή δυναμική αγορά εργασίας. Καθώς, η τεχνολογία *AI* συνεχίζει να εξελίσσεται, μπορούμε να περιμένουμε ακόμη πιο εξελιγμένα εργαλεία και τεχνικές για την ανάλυση των αγγελιών θέσεων εργασίας, που θα οδηγήσουν σε ένα πιο αποτελεσματικό, διαφανές και δίκαιο τοπίο προσλήψεων. (Mehra, 2020)

3.2.1. Η σημασία τις τεχνητής νοημοσύνης

Η τεχνητή νοημοσύνη (TN) αναφέρεται στην ικανότητα των μηχανών να μιμούνται τις ανθρώπινες γνωστικές λειτουργίες, όπως η μάθηση, η συλλογιστική και η επίλυση προβλημάτων. Δεν πρόκειται για τη δημιουργία ευαίσθητων ρομπότ, αλλά για την ανάπτυξη μηχανών που μπορούν να επεξεργάζονται πληροφορίες, να αναγνωρίζουν μοτίβα και να λαμβάνουν αποφάσεις με τρόπο που να προσομοιάζεται την ανθρώπινη νοημοσύνη. Ο τομέας αυτός περιλαμβάνει ένα ευρύ φάσμα τεχνολογιών, όπως η μηχανική μάθηση και η επεξεργασία φυσικής γλώσσας. Οι πιθανές εφαρμογές της TN είναι τεράστιες και εξακολουθούν να διερευνώνται. Ωστόσο, η άνοδος της τεχνητής νοημοσύνης εγείρει επίσης ερωτήματα σχετικά με τις ηθικές επιπτώσεις της. Πρέπει να εξετάσουμε ζητήματα μεροληψίας, εκτόπισης θέσεων εργασίας και το ενδεχόμενο κακής χρήσης. Καθώς, προχωράμε μπροστά, είναι ζωτικής σημασίας να αναπτύξουμε την τεχνητή νοημοσύνη με υπευθυνότητα, διασφαλίζοντας ότι θα ωφελήσει την ανθρωπότητα και θα προωθήσει ένα πιο δίκαιο μέλλον.

Η τεχνητή νοημοσύνη είναι μια τεχνολογία επιτρέπει στις εφαρμογές υπολογιστών να μαθαίνουν από την εμπειρία μέσω επαναληπτικής επεξεργασίας και αλγοριθμικής εκπαίδευσης. Τα συστήματα TN γίνονται πιο έξυπνα με κάθε επιτυχή επεξεργασία δεδομένων, καθώς κάθε αλληλεπίδραση επιτρέπει στο σύστημα να δοκιμάζει, να μετρά λύσεις και να αναπτύσσει τεχνογνωσία. Δεδομένου ότι αυτό μπορεί να ολοκληρωθεί γρήγορα, πολύ ταχύτερα από τον ρυθμό που ένας άνθρωπος θα ήταν σε θέση να εκτελέσει παρόμοια εργασία, Τα συστήματα TN μπορούν να γίνουν εμπειρογνώμονες πολύ ταχύτερα από τους ανθρώπους, καθιστώντας τα απίστευτα αποτελεσματικές επιλογές για οποιαδήποτε διαδικασία απαιτεί έξυπνη λήψη αποφάσεων. Αυτό καθιστά την τεχνητή νοημοσύνη μια απίστευτα ισχυρή και εξαιρετικά πολύτιμη τεχνολογία, αφού ουσιαστικά επιτρέπει στους υπολογιστές να σκέφτονται και να συμπεριφέρονται όπως οι άνθρωποι, αλλά σε πολύ ταχύτητες και με πολύ μεγαλύτερη επεξεργαστική ισχύ από ό,τι μπορεί να παράγει ο ανθρώπινος εγκέφαλος. Σε πολλές περιπτώσεις μια ποικιλία διαφορετικών εφαρμογών, τα συστήματα TN είναι ικανά να ξεπεράσουν σημαντικά τους ανθρώπους, και αυτός είναι ο κύριος λόγος για τον οποίο η τεχνολογία TN έχει γίνει τόσο σημαντική για τη σύγχρονη οικονομία.

Η τεχνητή νοημοσύνη έχει φέρει την επανάσταση στον κόσμο και έχει αλλάξει το μέλλον σε κάθε κλάδο, με εξαιρετικά αποτελέσματα. Ακολουθούν μερικές μόνο από τις συγκεκριμένες εφαρμογές όπου η τεχνητή νοημοσύνη έχει αποδειχθεί απίστευτα αποτελεσματική:

- ✚ Λιανικό εμπόριο :Η τεχνολογία ΑΙ εφαρμόζεται σε περιβάλλοντα λιανικού εμπορίου για τη διαχείριση των αποθεμάτων, το σχεδιασμό αποτελεσματικότερων διαρρυθμίσεων των καταστημάτων και την παροχή εξατομικευμένων συστάσεων αγορών μέσω των προτάσεων "Μπορεί επίσης να σας αρέσει" της *Amazon*, καθώς και εξατομικευμένων προτάσεων προβολής μέσω του αλγορίθμου συστάσεων του *Netflix* που βασίζεται στη μηχανική μάθηση.
- ✚ Παραγωγή : Οι λύσεις τεχνητής νοημοσύνης χρησιμοποιούνται για την πρόβλεψη του φορτίου και της ζήτησης για τα εργοστάσια, διασφαλίζοντας την όσο το δυνατόν πιο αποδοτική λειτουργία τους. Βοηθώντας στη λήψη καλύτερων αποφάσεων σχετικά με την εφοδιαστική και τον προγραμματισμό για την παραγγελία υλικών, τα χρονοδιαγράμματα για την ολοκλήρωση του έργου κ.λπ.
- ✚ Τραπεζικές υπηρεσίες :Τα συστήματα ΑΙ χρησιμοποιούνται για την επανεξέταση των χρηματοοικονομικών συναλλαγών για τον εντοπισμό δόλιων δραστηριοτήτων, την αξιολόγηση της πιστωτικής βαθμολογίας με μεγαλύτερη ακρίβεια και την αυτοματοποίηση εργασιών που απαιτούν χειροκίνητη εισαγωγή δεδομένων και διαχείριση δεδομένων.
- ✚ Βιοεπιστήμες: Η τεχνολογία ΤΝ εφαρμόζεται για τη δοκιμή νέων φαρμάκων, επιτρέποντας στους οργανισμούς να τα διαθέσουν ταχύτερα στην αγορά, και για την ανάλυση μεγάλων και πολύπλοκων συνόλων δεδομένων που βοηθούν στην ανακάλυψη νέων, αποτελεσματικότερων θεραπειών και φαρμακευτικών φαρμάκων. (CSU GLOBAL, 2021)

Η ΤΝ παρέχει σε ένα πρόγραμμα υπολογιστή με την δυνατότητα να σκέφτεται και να μαθαίνει μόνο του. Είναι μια προσομοίωση της ανθρώπινης νοημοσύνης σε μηχανές για να κάνουν πράγματα που κανονικά θα βασιζόμασταν στους ανθρώπους. Αυτό το τεχνολογικό θαύμα εκτείνεται πέρα από την απλή αυτοματοποίηση, ενσωματώνοντας ένα ευρύ φάσμα δεξιοτήτων τεχνητής νοημοσύνης και ικανοτήτων. Επιτρέπουν στις μηχανές να κατανοούν, να σκέφτονται, να μαθαίνουν και να αλληλοεπιδρούν με τρόπο που μοιάζει με τον ανθρώπινο. Υπάρχουν τρεις βασικοί τύποι τεχνητής νοημοσύνης με βάση τις δυνατότητές τους :

- a. Τεχνητή Στενή Νοημοσύνη (ΤΣΝ):Εστιάζει στην εκτέλεση μίας συγκεκριμένης εργασίας ή μίας κατηγορίας εργασιών. Καθώς, λειτουργεί με βάση

προκαθορισμένους κανόνες και αλγόριθμους. Δεν έχει την ικανότητα να μαθαίνει ή να προσαρμόζεται από μόνη της. Παραδείγματος χάριν, τα συστήματα αναγνώρισης προσώπου, αυτόματες μεταφράσεις και chatbots.

- b. Τεχνητή Γενική Νοημοσύνη (ΤΓΝ): Αποτελεί ένα θεωρητικό είδος ΤΝ έχει αντίστοιχη νοημοσύνη με έναν άνθρωπο. Θα μπορούσε να μαθαίνει, να κατανοεί και να εκτελεί οποιοδήποτε είδος εργασίας όμως δεν έχει ακόμη υλοποιηθεί πλήρως.
- c. Τεχνητή Υπερ-Νοημοσύνη (ΤΥΝ): Είναι ένα υποθετικό είδος τεχνητής νοημοσύνης που ξεπερνάει την ανθρώπινη νοημοσύνη και μπορεί να εκτελέσει οποιαδήποτε εργασία καλύτερα από τον άνθρωπο. Καθώς, θα μπορέσει να λύσει σύνθετα προβλήματα όμως, ενέχει ηθικές και φιλοσοφικές ανησυχίες. Ένα παράδειγμα, θα μπορούσε να είναι οι υπερ-ευφυείς μηχανές που θα μπορούσαν να εξελιχθούν ανεξάρτητα και να θέσουν σε κίνδυνο την ανθρωπότητα.

Κάθε τύπος ΤΝ έχει τα δικά του πλεονεκτήματα και μειονεκτήματα. Η ΤΣΝ είναι χρήσιμη για αυτοματοποιημένες εργασίες ενώ η ΤΥΝ εγείρει σοβαρές ανησυχίες. Η ανάπτυξη και η χρήση της ΤΝ φέρνει ηθικά διλήμματα καθώς, απαιτεί προσεκτική εξέταση. (simplilearn, 2024)

3.2.2. Πλεονεκτήματα και Μειονεκτήματα της Τεχνητής Νοημοσύνης

Η τεχνητή νοημοσύνη (AI) έχει γίνει μια μετασχηματιστική δύναμη στον κόσμο μας, αυτό μπορεί να περιλαμβάνει την αναγνώριση προτύπων, την επίλυση προβλημάτων, τη λήψη αποφάσεων και πολλά άλλα. Με τη χρήση αλγορίθμων, μοντέλων μηχανικής μάθησης και νευρωνικών δικτύων, η τεχνητή νοημοσύνη έχει προσφέρει λύσεις σε προβλήματα που παραδοσιακά θεωρούνταν δύσκολα για τους υπολογιστές. Ενώ, οι δυνατότητές της τεχνητής νοημοσύνης εγείρει ανησυχίες σχετικά με τα πιθανά μειονεκτημά της. Ας εμβαθύνουμε τόσο στα πλεονεκτήματα όσο και στα μειονεκτήματα της ΤΝ για να αποκτήσουμε μια ισορροπημένη προοπτική. Τα πλεονεκτήματα που προσφέρει η τεχνητή νοημοσύνη είναι τα ακόλουθα:

- ✚ Μείωση του ανθρώπινου σφάλματος: Ένα από τα μεγαλύτερα οφέλη της τεχνητής νοημοσύνης είναι η μείωση των σημαντικών λαθών και η αύξηση ακρίβειας. Οι αποφάσεις που λαμβάνει η ΤΝ σε κάθε βήμα αποφασίζονται από πληροφορίες που έχουν συλλεχθεί προηγουμένως και από ένα συγκεκριμένο σύνολο αλγορίθμων. Όταν προγραμματίζονται σωστά, τα σφάλματα αυτά μπορούν να μηδενιστούν.
- ✚ Μηδενικοί κίνδυνοι: Ένα άλλο μεγάλο πλεονέκτημα της τεχνητής νοημοσύνης είναι ότι οι άνθρωποι μπορούν να ξεπεράσουν πολλούς κινδύνους αφήνοντας τα ρομπότ τεχνητής νοημοσύνης να κάνουν για εμάς αυτές τις διαδικασίες. Επιπλέον, μπορούν να παρέχουν ακριβή εργασία με μεγαλύτερη υπευθυνότητα. Ένα παράδειγμα, μηδενικών κινδύνων είναι μια πλήρως αυτοματοποιημένη γραμμή παραγωγής σε μια μονάδα παραγωγής. Τα ρομπότ εκτελούν όλες τις εργασίες, εξαλείφοντας τον κίνδυνο ανθρώπινου λάθους και τραυματισμού σε επικίνδυνα περιβάλλοντα. (simplilearn, 2024)
- ✚ Βελτιωμένη απόδοση και αυτοματισμός: Ένα από τα πιο σημαντικά πλεονεκτήματα της ΑΙ είναι η ικανότητά του να αυτοματοποιεί επαναλαμβανόμενες εργασίες. Οι αλγόριθμοι ΑΙ μπορούν να αναλύσουν τεράστιες ποσότητες δεδομένων και να εκτελέσουν εργασίες με απaráμιλλη ταχύτητα και ακρίβεια. Αυτό απελευθερώνει ανθρώπινο δυναμικό για πιο σύνθετες και δημιουργικές προσπάθειες. Τα ρομπότ με τεχνητή νοημοσύνη εκτελούν κουραστικές εργασίες με μεγαλύτερη ακρίβεια και αποτελεσματικότητα, ενισχύοντας την παραγωγικότητα και μειώνοντας τα σφάλματα. Τα *Chatbots* που υποστηρίζονται από ΑΙ μπορούν να απαντούν σε συχνές ερωτήσεις, να χειρίζονται βασικές ερωτήσεις πελατών και να προγραμματίζουν ραντεβού, ελευθερώνοντας ανθρώπινους πράκτορες για πιο περίπλοκα ζητήματα. Επι προσθέτως, οι αλγόριθμοι τεχνητής νοημοσύνης μπορούν να αναλύσουν τεράστια σύνολα δεδομένων, να αυτοματοποιήσουν την εισαγωγή δεδομένων και να δημιουργήσουν αναφορές, μειώνοντας σημαντικά τον χρόνο επεξεργασίας και τον ανθρώπινο φόρτο εργασίας. (Russell, 2016)
- ✚ Βελτιωμένη λήψη αποφάσεων: Η τεχνητή νοημοσύνη υπερέχει στην ανάλυση τεράστιων ποσοτήτων δεδομένων για τον εντοπισμό μοτίβων και τάσεων που μπορεί να χάσουν οι άνθρωποι. Αυτή η ικανότητα εξουσιοδοτεί την καλύτερη λήψη αποφάσεων σε διάφορους τομείς. Οι αλγόριθμοι τεχνητής νοημοσύνης μπορούν να

αναλύσουν οικονομικά δεδομένα για να προβλέψουν τις τάσεις της αγοράς, να αξιολογήσουν την πιστοληπτική ικανότητα και να εξατομικεύσουν επενδυτικές στρατηγικές. Η τεχνητή νοημοσύνη μπορεί να αναλύσει ιστορικά δεδομένα για να προβλέψει πιθανούς κινδύνους και να εφαρμόσει προληπτικά μέτρα, βελτιώνοντας την ασφάλεια σε διάφορους τομείς όπως η χρηματοδότηση.

- ✚ Καινοτομία και Επιστημονική Ανακάλυψη: Η τεχνητή νοημοσύνη δεν αφορά μόνο την αυτοματοποίηση έχει επίσης τεράστιες δυνατότητες για καινοτομία και επιστημονική ανακάλυψη. Τα εργαλεία που λειτουργούν με τεχνητή νοημοσύνη μπορούν να βοηθήσουν τους ερευνητές να αναλύσουν πολύπλοκα επιστημονικά σύνολα δεδομένων, να αποκαλύψουν νέα μοτίβα και σχέσεις .
- ✚ Έρευνα και Ανάλυση Δεδομένων: Με τη βοήθεια της τεχνητής νοημοσύνης, οι επιστήμονες της έρευνας και των δεδομένων είναι σε θέση να αναλύουν καλύτερα μοτίβα, να προβλέψουν τα αποτελέσματα και να κάνουν γρήγορες προσαρμογές. Οι πληροφορίες που θα χρειάζονταν μήνες για να συλλεχθούν τώρα με ταχύτατους ρυθμούς. (WGU, 2022)
- ✚ Αυτοματοποίηση απόκρισης απειλών: Ένα άλλο πλεονέκτημα της *AI* είναι η ικανότητά του να αυτοματοποιεί την απόκριση σε απειλές. Οι αλγόριθμοι μπορούν να προγραμματιστούν ώστε να ανταποκρίνονται αυτόματα σε συμβάντα ασφαλείας, ενεργοποιώντας ανταποκρίνονται ταχύτερα και αποτελεσματικότερα. Για παράδειγμα, εάν ένα σύστημα *AI* μπορεί να εμποδίσει αυτόματα την πρόσβαση στο παραβιασμένο σύστημα, αποτρέποντας περαιτέρω ζημιές. Μπορεί επίσης να στείλει ειδοποιήσεις στο προσωπικό ασφαλείας, παρέχοντάς τους πληροφορίες σε πραγματικό χρόνο για το συμβάν και δίνοντάς τους τη δυνατότητα να ανταποκριθούν γρήγορα. Επιπλέον, το *AI* μπορεί επίσης να αυτοματοποιήσει τη διαδικασία επιδιόρθωσης ευπαθειών και ενημέρωσης των ελέγχων ασφαλείας. Αυτό μπορεί να βοηθήσει στη μείωση του χρόνου που απαιτείται για την εφαρμογή ενημερώσεων ασφαλείας και να διασφαλίσει ότι τα συστήματα είναι πάντα ενημερωμένο και προστατευμένο από τις τελευταίες απειλές. Επίσης, η τεχνητή νοημοσύνη μπορεί να βοηθήσει τους οργανισμούς να ανταποκριθούν πιο γρήγορα και αποτελεσματικά σε συμβάντα ασφαλείας, μειώνοντας τον κίνδυνο παραβίασης δεδομένων. Συνολικά, η

αυτοματοποίηση της απόκρισης απειλών είναι ένα σημαντικό πλεονέκτημα της τεχνητής νοημοσύνης στην ασφάλεια των πληροφοριών.

- ✚ **Πραγματικού Χρόνου Εισαγωγές σε Συμβάντα Ασφαλείας:** Ένα άλλο πλεονέκτημα του *AI* είναι η ικανότητά του να παρέχει πληροφορίες σε πραγματικό χρόνο. Καθώς, μπορούν να αναλύσουν τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο, παρέχοντας οργανώσεις με άμεσες γνώσεις για συμβάντα και απειλές. Η τεχνητή νοημοσύνη μπορεί να επιτρέψει στους οργανισμούς να ανταποκριθούν γρήγορα και αποτελεσματικά σε πιθανές απειλές. Παρέχοντας στους οργανισμούς ανίχνευση απειλών σε πραγματικό χρόνο και τις δυνατότητες απόκρισης, η τεχνητή νοημοσύνη μπορεί να βοηθήσει τους οργανισμούς να παραμείνουν μπροστά από τις εξελισσόμενες απειλές. (Mughal, 2018)

Η τεχνητή νοημοσύνη (*AI*) έχει φέρει επανάσταση καθώς παρουσιάζει αναμφισβήτητες προόδους. Εκτός, από τα οφέλη που μπορούμε να αποκομίσουμε από την τεχνητή νοημοσύνη παρουσιάζει επίσης, μια σειρά από πιθανά μειονεκτήματα που απαιτούν προσεκτική εξέταση. Σε αυτήν την διπλωματική θα εμβαθύνουμε σε μερικά από τα πιο ανησυχητικά μειονεκτήματα της τεχνητής νοημοσύνης. Αυτά είναι:

- ✚ **Η μετατόπιση εργασίας και το μέλλον της εργασίας:** Μία από τις πιο ευρέως ανησυχίες σχετικά με την τεχνητή νοημοσύνη είναι η δυνατότητά της να αυτοματοποιεί εργασίες που εκτελούνται αυτήν τη στιγμή από ανθρώπους. Οι εξελίξεις της τεχνητής νοημοσύνης σε διάφορους κλάδους, από την κατασκευή έως την εξυπηρέτηση πελατών, απειλούν να εκτοπίσουν τους εργαζομένους των οποίων οι εργασίες μπορούν να αναπαραχθούν από μηχανές. Αυτό εγείρει σημαντικές ανησυχίες σχετικά με τα ποσοστά ανεργίας, την εισοδηματική ανισότητα και την ανάγκη για προγράμματα επανεκπαίδευσης του εργατικού δυναμικού για να εξοπλίσουν τα άτομα με δεξιότητες σχετικές με την εξελισσόμενη αγορά εργασίας. (Carl Benedikt Frey, 2017)

- ✚ **Δαπανηρή υλοποίηση:** Το μεγαλύτερο και πιο προφανές μειονέκτημα της εφαρμογής του *AI* είναι ότι η ανάπτυξή του μπορεί να είναι εξαιρετικά δαπανηρή. Το κόστος για μια πλήρως εφαρμοσμένη λύση τεχνητής νοημοσύνης για τις περισσότερες

επιχειρήσεις κυμαινόταν από 20.000\$ έως αρκετά εκατομμύρια .Καθώς, με την πλήρως εφαρμογής της τεχνητή νοημοσύνη μπορεί να βοηθήσει στον εξορθολογισμό της ροής εργασίας.

- ✚ Έλλειψη δημιουργικότητα: Ένα μεγάλο μειονέκτημα της τεχνητής νοημοσύνης είναι ότι δεν μπορεί να μάθει να σκέφτεται έξω από το κουτί. Η τεχνητή νοημοσύνη είναι ικανή να μαθαίνει με την πάροδο του χρόνου με προ-τροφοδοτημένα δεδομένα και προηγούμενες εμπειρίες, αλλά δεν μπορεί να είναι δημιουργική στην προσέγγισή της.
- ✚ Ηθικά ζητήματα: Η ηθική είναι ένα από τα σημαντικά ανθρώπινα χαρακτηριστικά που μπορεί να είναι δύσκολο να ενσωματωθούν σε μια τεχνητή νοημοσύνη. Η ταχεία πρόοδος της τεχνητής νοημοσύνης έχει εγείρει μια σειρά από ανησυχίες ότι, η τεχνητή νοημοσύνη θα αναπτυχθεί ανεξέλεγκτα και τελικά θα εξαφανίσει την ανθρωπότητα. (simplilearn, 2024)
- ✚ Αλγοριθμική προκατάληψη και δικαιοσύνη: Οι αλγόριθμοι τεχνητής νοημοσύνης είναι κατάλληλη για δεδομένα στα οποία έχουν εκπαιδευτεί. Εάν τα δεδομένα εκπαίδευσης είναι προκατειλημμένα, τα μοντέλα τεχνητής νοημοσύνης που προκύπτουν μπορούν να ενισχύσουν τις κοινωνικές ανισότητες. Αυτό μπορεί να οδηγήσει σε πρακτικές πρόσληψης που εισάγουν διακρίσεις, αρνήσεις δανείων ή μεροληπτικά αποτελέσματα ποινικής δικαιοσύνης. Ο μετριασμός της αλγοριθμικής μεροληψίας απαιτεί προσεκτική επιλογή δεδομένων, διαφορετικά σύνολα εκπαίδευσης και συνεχή παρακολούθηση των διαδικασιών λήψης αποφάσεων.
- ✚ Έλλειψη επεξηγηματικότητας και διαφάνειας: Πολλά συστήματα τεχνητής νοημοσύνης, ιδιαίτερα αυτά που βασίζονται στη βαθιά μάθηση, μπορεί να είναι πολύπλοκα «μαύρα κουτιά». Μπορεί να είναι δύσκολο να κατανοήσουμε πώς ένα μοντέλο AI καταλήγει σε μια συγκεκριμένη απόφαση. Αυτή η έλλειψη διαφάνειας μπορεί να εγείρει ανησυχίες .Ιδιαίτερα όταν η τεχνητή νοημοσύνη χρησιμοποιείται σε εφαρμογές υψηλού κινδύνου, όπως οικονομικές αποφάσεις ή αυτόνομα οχήματα.

Η τεχνητή νοημοσύνη αποτελεί ένα ισχυρό εργαλείο με τη δυνατότητα να αντιμετωπίσει μερικές από τις πιο πιεστικές προκλήσεις της ανθρωπότητας και να βελτιώσει τη ζωή μας με βαθύτατους τρόπους. Καθώς προχωράμε, η προώθηση της υπεύθυνης ανάπτυξης, η προτεραιότητα της συνεργασίας ανθρώπου-τεχνητής νοημοσύνης και η αντιμετώπιση ηθικών ανησυχιών θα είναι ουσιαστικής σημασίας για να διασφαλιστεί ότι η τεχνητή νοημοσύνη λειτουργεί ως δύναμη.

3.2.3. Η τεχνητή νοημοσύνης διευκολύνει την εξαγωγή πληροφορίας από διαδικτυακές αγγελίες εργασίας.

Στη σημερινή ψηφιακή εποχή, ο πολλαπλασιασμός των διαδικτυακών αγγελιών εργασίας έχει μεταμορφώσει σημαντικά το τοπίο των προσλήψεων, προσφέροντας ένα ευρύ φάσμα ευκαιριών τόσο για όσους αναζητούν εργασία όσο και για τους εργοδότες. Ωστόσο, μέσα σε αυτήν την αφθονία πληροφοριών, η εξαγωγή σχετικών πληροφοριών από τις αναρτήσεις εργασίας μπορεί να είναι μια αποθαρρυντική εργασία. Ευτυχώς, οι εξελίξεις στην τεχνητή νοημοσύνη έχουν φέρει επανάσταση στη διαδικασία, καθιστώντας ευκολότερη την αναζήτηση τεράστιων ποσοτήτων δεδομένων, την εξαγωγή σχετικών πληροφοριών και την απόκτηση πολύτιμων πληροφοριών.

Η Τεχνητή Νοημοσύνη και ιδιαίτερα οι αλγόριθμοι Επεξεργασίας Φυσικής Γλώσσας, διαδραματίζει καθοριστικό ρόλο στην αποκρυπτογράφηση της πολυπλοκότητας της γλώσσας στις αγγελίες εργασίας. Οι αλγόριθμοι *NLP* είναι ικανοί στην κατανόηση και ανάλυση δεδομένων κειμένου, επιτρέποντάς τους να εξάγουν βασικές πληροφορίες όπως απαιτήσεις εργασίας, προσόντα, ευθύνες και εταιρικές προτιμήσεις. Αξιοποιώντας τεχνικές όπως η αναγνώριση επώνυμων οντοτήτων, η ανάλυση συναισθήματος και η εξαγωγή λέξεων-κλειδίων. Τα συστήματα τεχνητής νοημοσύνης μπορούν να περιηγηθούν γρήγορα στις αναρτήσεις θέσεων εργασίας, εντοπίζοντας φράσεις-κλειδιά και όρους που σχετίζονται τόσο με όσους αναζητούν εργασία όσο και με τους εργοδότες. Επιπλέον, οι τεχνικές εξαγωγής δεδομένων που βασίζονται στην *AI* επιτρέπουν την αυτόματη ανάλυση δομημένων πληροφοριών από μη δομημένες αναρτήσεις εργασίας. Πολλές διαδικτυακές λίστες θέσεων εργασίας ακολουθούν μια τυποποιημένη μορφή, συμπεριλαμβανομένων πεδίων όπως ο τίτλος εργασίας, όνομα εταιρείας, τοποθεσία, μισθός και απαιτούμενα προσόντα. Ωστόσο, η μη αυτόματη εξαγωγή αυτών των δομημένων δεδομένων από διαφορετικές πηγές μπορεί να είναι χρονοβόρα και επιρρεπής σε σφάλματα. Οι αλγόριθμοι τεχνητής νοημοσύνης μπορούν αυτόματα να

εντοπίζουν και να εξάγουν σχετικές πληροφορίες από αγγελίες εργασίας, μετατρέποντας το μη δομημένο κείμενο σε δομημένα δεδομένα που μπορούν εύκολα να υποβληθούν σε επεξεργασία και ανάλυση . Αυτή η αυτοματοποίηση μειώνει σημαντικά την επιβάρυνση των ατόμων που αναζητούν εργασία, δίνοντάς τους τη δυνατότητα να εντοπίζουν γρήγορα ευκαιρίες εργασίας που ταιριάζουν με τις προτιμήσεις και τα προσόντα τους.

Οι αλγόριθμοι αντιστοίχισης εργασίας με τεχνητή νοημοσύνη ενισχύουν την αποδοτικότητα και την αποτελεσματικότητα της διαδικασίας αναζήτησης εργασίας παρέχοντας εξατομικευμένες προτάσεις στους υποψηφίους. Αυτοί οι αλγόριθμοι αναλύουν διάφορους παράγοντες, συμπεριλαμβανομένων των δεξιοτήτων, της εμπειρίας, των προτιμήσεων και των απαιτήσεων που καθορίζονται στις αγγελίες εργασίας του υποψηφίου, για να εντοπίσουν τις κατάλληλες αντιστοιχίσεις. Αξιοποιώντας τεχνικές μηχανικής μάθησης, όπως το συνεργατικό φιλτράρισμα και το φιλτράρισμα βάσει περιεχομένου, τα συστήματα *AI* μπορούν να προτείνουν σχετικές ευκαιρίες εργασίας προσαρμοσμένες στο προφίλ κάθε υποψηφίου .Αυτή η εξατομικευμένη προσέγγιση όχι μόνο εξοικονομεί χρόνο για όσους αναζητούν εργασία, αλλά αυξάνει επίσης, την πιθανότητα εύρεσης εργασίας που να ευθυγραμμίζεται με τα προσόντα και τους στόχους της σταδιοδρομίας τους.

Εκτός από το όφελος των ατόμων που αναζητούν εργασία, η εξαγωγή πληροφοριών με γνώμονα την τεχνητή νοημοσύνη από διαδικτυακές αναρτήσεις θέσεων εργασίας προσφέρει επίσης πλεονεκτήματα στους εργοδότες. Με την αυτοματοποίηση της διαδικασίας ελέγχου και ανάλυσης των αιτήσεων εργασίας, οι αλγόριθμοι τεχνητής νοημοσύνης επιτρέπουν στους υπεύθυνους προσλήψεων να εντοπίζουν τους κορυφαίους υποψηφίους πιο αποτελεσματικά. Για παράδειγμα, τα συστήματα αυτά μπορούν να αντιστοιχίσουν αυτόματα τα βιογραφικά των υποψηφίων με τις απαιτήσεις εργασίας, να αξιολογήσουν τα προσόντα των υποψηφίων και ακόμη και να πραγματοποιήσουν αρχικές εξετάσεις μέσω chatbots ή εικονικών βοηθών. Αυτή η αυτοματοποίηση μειώνει τον χρόνο και τους πόρους που απαιτούνται για τη διαδικασία πρόσληψης, επιτρέποντας στους οργανισμούς να καλύψουν τις κενές θέσεις γρηγορότερα και πιο αποτελεσματικά.

Συμπερασματικά, η τεχνητή νοημοσύνη φέρνει επανάσταση στον τρόπο εξαγωγής πληροφοριών από τις διαδικτυακές αναρτήσεις εργασίας, καθιστώντας τη διαδικασία αναζήτησης εργασίας ευκολότερη και πιο αποτελεσματική τόσο για όσους αναζητούν εργασία όσο και για εκείνους που προσλαμβάνουν προσωπικό για εργασία. Αξιοποιώντας τεχνικές *NLP*, αλγόριθμους εξαγωγής δεδομένων και εξατομικευμένους αλγόριθμους αντιστοίχισης εργασιών, η τεχνητή νοημοσύνη επιτρέπει την αυτόματη ανάλυση των αναρτήσεων θέσεων εργασίας, τον εντοπισμό σχετικών ευκαιριών και την αντιστοίχιση υποψηφίων με κατάλληλες

θέσεις. Καθώς, η τεχνητή νοημοσύνη συνεχίζει να προοδεύει, στο τοπίο των προσλήψεων ενισχύοντας την αποτελεσματικότητα και την καινοτομία στην αγορά εργασίας.

ΚΕΦΑΛΑΙΟ 4: Η ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΚΑΙ ΤΕΧΝΙΚΩΝ ΕΞΑΓΩΓΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

4.1. Η έννοια της φυσικής γλώσσας επεξεργασίας (NLP) και η εφαρμογή στην αφαίρεση πληροφορίας

Η επεξεργασία φυσικής γλώσσας (*Natural Language Processing*), σχετίζεται με την τεχνητή νοημοσύνη και την επιστήμη των υπολογιστών. Όμως, οι υπολογιστές θα πρέπει να είναι σε θέση να κατανοούν και να χειρίζονται την ανθρώπινη γλώσσα, σε όλες τις γραπτές, προφορικές, ακόμη και υπογεγραμμένες μορφές της. Το *NLP* ενώνει το χάσμα επικοινωνίας μεταξύ ανθρώπων και μηχανών, επιτρέποντάς μας να αλληλοεπιδρούμε με τους υπολογιστές με πιο φυσικό τρόπο.

Ως επεξεργασία φυσικής γλώσσας (*NLP*) ορίζουμε την ικανότητα ενός προγράμματος υπολογιστή να κατανοεί την ανθρώπινη γλώσσα όπως προφέρεται και γράφεται που αναφέρεται ως φυσική γλώσσα. Καθώς, αποτελεί ένα συστατικό της τεχνητής νοημοσύνης . Το *NLP* υπάρχει για περισσότερα από 50 χρόνια και προέρχεται από τον χώρο της γλωσσολογίας. Έχει μια ποικιλία πραγματικών εφαρμογών σε πολλούς τομείς, συμπεριλαμβανομένης της ιατρικής έρευνας, των μηχανών αναζήτησης και της επιχειρηματικής ευφυΐας. (Alexander S. Gillis, n.d.) Χρησιμοποιεί σε βασισμένες κανόνες και σε προσεγγίσεις μηχανικής μάθησης για την κατανόηση της δομής και του νοήματος του κειμένου. Όπως, είναι τα chatbots, τους βοηθούς φωνής, τα προγράμματα σάρωσης που βασίζονται σε κείμενο, οι εφαρμογές μετάφρασης και το εταιρικό λογισμικό βοηθά στις επιχειρηματικές λειτουργίες, αυξάνει την παραγωγικότητα και απλοποιεί διαφορετικές διαδικασίες. (Staff, 2024) Η επεξεργασία φυσικής γλώσσας (*Natural Language Processing - NLP*) λειτουργεί μέσω ενός συνδυασμού αλγορίθμων, μοντέλων μηχανικής μάθησης και γλωσσικών κανόνων που επιτρέπουν στους υπολογιστές να κατανοήσουν, να αναλύσουν και να παράγουν φυσική γλώσσα σε διάφορες μορφές.

4.1.1. Εξαγωγή πληροφορίας με την χρήση των τεχνικών *NLP*

Η εξόρυξη κειμένου, είναι γνωστή και ως εξόρυξη δεδομένων κειμένου ή ανακάλυψη γνώσης από βάσεις δεδομένων κειμένου, περιλαμβάνει τη διαδικασία άντλησης πληροφοριών υψηλής ποιότητας από κείμενο. Συνήθως περιλαμβάνει εργασίες όπως η κατηγοριοποίηση κειμένου, η ομαδοποίηση κειμένου, η εξαγωγή κανόνων συσχέτισης και η οπτικοποίηση κειμένου. Η θεμελιώδης δομή της εξόρυξης κειμένου αποτελείται από δύο βασικά στάδια. Αυτά είναι ο εξευγενισμός κειμένου και την απόσταξη γνώσης. Ο εξευγενισμός κειμένου μετατρέπει έγγραφα κειμένου ελεύθερης μορφής σε μια ενδιάμεση μορφή, ενώ η απόσταξη γνώσης αντλεί πρότυπα ή γνώση από αυτή την ενδιάμεση μορφή.

Οι βασικές τεχνικές εξόρυξης κειμένου περιλαμβάνουν:

1. Κατηγοριοποίηση κειμένου: Περιλαμβάνει την ανάθεση προκαθορισμένων κατηγοριών σε έγγραφα με βάση το περιεχόμενό τους. Χρησιμοποιείται συνήθως στην οργάνωση μεγάλων συνόλων δεδομένων σε πιο εύχρηστες υποομάδες.
2. Ομαδοποίηση κειμένου: Αυτή η τεχνική ομαδοποιεί παρόμοια έγγραφα σε ομάδες με βάση το περιεχόμενό τους, βοηθώντας στον εντοπισμό μοτίβων και τάσεων εντός των δεδομένων.
3. Εξαγωγή κανόνων συσχέτισης: Αυτή η μέθοδος ανακαλύπτει ενδιαφέρουσες σχέσεις μεταξύ μεταβλητών σε μεγάλα σύνολα δεδομένων, που χρησιμοποιούνται συχνά για την ανάλυση καλαθιού αγοράς.
4. Οπτικοποίηση κειμένου: Περιλαμβάνει τη γραφική αναπαράσταση δεδομένων κειμένου για να βοηθήσει στην κατανόηση σύνθετων σχέσεων εντός των δεδομένων.

Η εξαγωγή πληροφοριών (*IE*) είναι μια κρίσιμη εργασία *NLP* που επικεντρώνεται στον εντοπισμό συγκεκριμένων πληροφοριών μέσα σε ένα κείμενο. Η διαδικασία αυτή περιλαμβάνει την αναγνώριση βασικών φράσεων, οντοτήτων και σχέσεων μέσα σε ένα έγγραφο. Η *IE* περιλαμβάνει συνήθως την αντιστοίχιση προτύπων, όπου χρησιμοποιούνται προκαθορισμένα πρότυπα ή μοτίβα για τον εντοπισμό σχετικών πληροφοριών. Η *NLP* και η εξόρυξη κειμένου έχουν εφαρμοστεί σε διάφορους τομείς, αποδεικνύοντας την ευελιξία και την αποτελεσματικότητά τους. Για παράδειγμα, στον τομέα της ακαδημαϊκής έρευνας, οι τεχνικές αυτές έχουν χρησιμοποιηθεί για την ανάλυση τάσεων και μοτίβων σε επιστημονικά άρθρα.

Παρά τις προόδους στον τομέα του *NLP* και της εξόρυξης κειμένου, παραμένουν αρκετές προκλήσεις. Η μη δομημένη φύση των δεδομένων κειμένου απαιτεί συχνά εκτεταμένη προ επεξεργασία, συμπεριλαμβανομένης της δημιουργίας συμβόλων (*tokenization*), της

αφαίρεσης των σταματημένων λέξεων και της κανονικοποίησης. Επιπλέον, η πολυπλοκότητα της φυσικής γλώσσας, με τις αποχρώσεις και τις ασάφειές της, θέτει σημαντικά εμπόδια για τα αυτοματοποιημένα συστήματα. Η μελλοντική έρευνα κατευθύνεται προς τη βελτίωση της ακρίβειας και της αποτελεσματικότητας των αλγορίθμων *NLP*, των ενσωμάτωση δεδομένων (όπως κείμενο, εικόνες και γραφήματα) και την ανάπτυξη πιο εξελιγμένων μεθόδων για την κατανόηση του πλαισίου και της σημασιολογίας του κειμένου. Επιπλέον, αναμένεται να αυξηθεί η τάση της διατομεακής έρευνας, όπου τα ευρήματα από έναν τομέα επηρεάζουν άλλους τομείς, ενισχύοντας περαιτέρω τη δυνατότητα εφαρμογής των τεχνικών εξόρυξης κειμένου και *NLP*.

Η *NLP* και η εξόρυξη κειμένου είναι ισχυρά εργαλεία για την εξαγωγή πληροφοριών από μεγάλους όγκους αδόμητου κειμένου. Μέσω τεχνικών όπως η κατηγοριοποίηση κειμένου, η ομαδοποίηση και η εξαγωγή πληροφοριών, οι μεθοδολογίες αυτές παρέχουν πολύτιμες γνώσεις σε διάφορους τομείς. Αν και οι προκλήσεις παραμένουν, η συνεχιζόμενη έρευνα και οι τεχνολογικές εξελίξεις υπόσχονται να ενισχύσουν τις δυνατότητες και τις εφαρμογές αυτών των τεχνικών, καθιστώντας απαραίτητες στην εποχή των μεγάλων δεδομένων και της ψηφιακής πληροφορίας. (Said A. Salloum, 2017)

4.2. Η ανάλυση μεθοδολογιών Machine Learning στην εξαγωγή πληροφορίας

Η εξαγωγή πληροφοριών (*IE*) είναι μια κρίσιμη εργασία στην επεξεργασία φυσικής γλώσσας (*NLP*), με στόχο την αυτόματη ανάκτηση συγκεκριμένων πληροφοριών από μη δομημένα δεδομένα. Οι μεθοδολογίες μηχανικής μάθησης έχουν βελτιώσει σημαντικά τις δυνατότητες των συστημάτων εξαγωγής πληροφορίας, διευκολύνοντας την εξαγωγή σχετικών δεδομένων από μεγάλα σώματα κειμένων με βελτιωμένη ακρίβεια και αποτελεσματικότητα. Παρακάτω παρατίθενται οι τεχνικές μηχανικής μάθησης στην εξαγωγή πληροφοριών. Αυτές είναι:

1. Μέθοδοι επιβλεπόμενης μάθησης: Η μάθηση με επίβλεψη περιλαμβάνει την εκπαίδευση ενός μοντέλου σε ένα σύνολο δεδομένων με ετικέτες, όπου τα ζεύγη εισόδου-εξόδου είναι γνωστά. Ο στόχος είναι η εκμάθηση μιας αντιστοίχισης από τις εισόδους στις εξόδους, ώστε να γίνουν προβλέψεις σε νέα δεδομένα.
2. Λογιστική παλινδρόμηση και μηχανές διανυσμάτων υποστήριξης (*SVM*): Αυτά τα γραμμικά μοντέλα χρησιμοποιούνται ευρέως για δυαδικές εργασίες ταξινόμησης στην εξαγωγή

πληροφοριών . Είναι αποτελεσματικά για εργασίες όπως η αναγνώριση ονομαστικών οντοτήτων (*NER*) και η ταξινόμηση κειμένου λόγω της απλότητας και της ερμηνευσιμότητάς τους.

3. Δέντρα αποφάσεων και τυχαία δάση: Τα δέντρα αποφάσεων είναι διαισθητικά μοντέλα που χωρίζουν τα δεδομένα σε υποσύνολα με βάση τις τιμές των χαρακτηριστικών, ενώ τα τυχαία δάση ενισχύουν την απόδοση με τη συγκέντρωση πολλαπλών δέντρων αποφάσεων. Αυτά τα μοντέλα είναι ανθεκτικά στην υπερπροσαρμογή και μπορούν να χειριστούν αποτελεσματικά μεγάλους χώρους χαρακτηριστικών. Στο πλαίσιο αυτό, χρησιμοποιούνται για εργασίες όπως η εξαγωγή σχέσεων και η ταξινόμηση εγγράφων.

4. Ταξινομητής *Naive Bayes*: Βασισμένος στο θεώρημα του *Bayes*, αυτός ο ταξινομητής πιθανότητας είναι ιδιαίτερα χρήσιμος για προβλήματα ταξινόμησης κειμένου. Υποθέτει ανεξαρτησία μεταξύ των χαρακτηριστικών, γεγονός που απλοποιεί τους υπολογισμούς αλλά ενδέχεται να περιορίσει την απόδοση σε πολύπλοκα σενάρια.

Οι μέθοδοι *Ensemble* συνδυάζουν προβλέψεις από πολλαπλά μοντέλα για να βελτιώσουν τη συνολική απόδοση. Τεχνικές όπως το *AdaBoost* και το *Gradient Boosting* είναι εξέχουσες σε αυτή την κατηγορία. Η μέθοδος *AdaBoost* επικεντρώνεται στη βελτίωση της ακρίβειας των αδύναμων ταξινομητών. Έχει εφαρμοστεί με επιτυχία σε εργασίες εξαγωγής πληροφορίας για τη βελτίωση της ακρίβειας και της ανάκλησης συνδυάζοντας αδύναμους εκπαιδευτές σε ένα ισχυρό προγνωστικό μοντέλο. Ενώ οι τεχνικές *Gradient Boosting* δημιουργεί μοντέλα διαδοχικά, όπου κάθε νέο μοντέλο διορθώνει τα σφάλματα των προηγούμενων. Είναι ιδιαίτερα αποτελεσματική στο χειρισμό σύνθετων προτύπων στα δεδομένα, γεγονός που την καθιστά κατάλληλη για περίπλοκες εργασίες εξαγωγής πληροφορίας, όπως η ταξινόμηση πολλαπλών κλάσεων και η εξαγωγή χαρακτηριστικών.

Οι μεθοδολογίες μηχανικής μάθησης έχουν φέρει επανάσταση στον τομέα της εξόρυξης πληροφοριών, παρέχοντας ισχυρά εργαλεία για την αυτοματοποίηση της ανάκτησης σχετικών δεδομένων από μεγάλα σώματα κειμένων. Η ενσωμάτωση της μηχανικής μάθησης και πιο συγκεκριμένα της επιβλεπόμενης μάθησης, έχει βελτιώσει την ακρίβεια και την αποτελεσματικότητα των συστημάτων εξαγωγής πληροφοριών . Η μελλοντική έρευνα θα πρέπει να επικεντρωθεί σε υβριδικά μοντέλα που συνδυάζουν τα πλεονεκτήματα διαφορετικών προσεγγίσεων *ML* και να διερευνήσει τη δυνατότητα εφαρμογής τους σε ποικίλες εργασίες και γλώσσες εξαγωγής πληροφορίας. (Kowsher, 2020)

4.3.Μελλοντικές προοπτικές για τη χρήση των LLMs στην εξαγωγή πληροφοριών

Η έλευση των μεγάλων γλωσσικών μοντέλων (*LLMs*), όπως το *ChatGPT*, το *GPT-3* και οι διάδοχοί τους, σηματοδότησε ένα σημαντικό άλμα στην επεξεργασία φυσικής γλώσσας (*NLP*). Αυτά τα μοντέλα όχι μόνο έχουν μεταμορφώσει τον τρόπο με τον οποίο αλληλοεπιδρούμε με την τεχνητή νοημοσύνη, αλλά έχουν επίσης ανοίξει νέους δρόμους για την εξαγωγή πληροφοριών σε διάφορους τομείς. Η παρούσα ανασκόπηση διερευνά τις μελλοντικές προοπτικές των *LLMs* στην εξαγωγή πληροφοριών, αντλώντας πληροφορίες από πρόσφατες μελέτες και τεχνολογικές εξελίξεις.

Οι *LLMs* έχουν εξελιχθεί από απλές τεχνικές ενσωμάτωσης λέξεων σε εξελιγμένες αρχιτεκτονικές μετασχηματιστών. Τα πρώιμα μοντέλα όπως το *Bag-of-Words (BoW)* και το *Latent Dirichlet Allocation (LDA)* άνοιξαν το δρόμο για σύγχρονες προσεγγίσεις όπως το *BERT (Bidirectional Encoder Representations from Transformers)* και το *GPT (Generative Pre-trained Transformers)*. Αυτές οι εξελίξεις επέτρεψαν στους *LLMs* να κατανοούν και να παράγουν κείμενο που μοιάζει με ανθρώπινο κείμενο, ενισχύοντας σημαντικά την ικανότητά τους να εξάγουν πληροφορίες από τεράστια και αδόμητα σύνολα δεδομένων. Οι δυνατότητες των *LLMs* έχουν επεκταθεί με την εισαγωγή μοντέλων όπως τα *GPT-3* και *GPT-4*, τα οποία αξιοποιούν εκτεταμένα σύνολα δεδομένων για να εκτελούν ένα ευρύ φάσμα εργασιών με βάση προτροπές φυσικής γλώσσας. Αυτή η ευελιξία τα καθιστά κατάλληλα για εφαρμογές σε τομείς τόσο διαφορετικούς όσο η υγειονομική περίθαλψη, η χρηματοδότηση και η ανάλυση των μέσων κοινωνικής δικτύωσης. Οι *LLMs* χρησιμοποιούνται ήδη σε διάφορες εργασίες εξόρυξης πληροφοριών. Για παράδειγμα, στο οικονομικό κλάδο οι *LLMs* χρησιμοποιούνται για την ανάλυση των τάσεων της αγοράς και τη δημιουργία οικονομικών αναφορών. Μια αξιοσημείωτη εφαρμογή είναι η χρήση των *LLMs* για την αναγνώριση ονομαστικών οντοτήτων (*Named Entity Recognition - NER*) και τη μοντελοποίηση θεμάτων, οι οποίες είναι κρίσιμες για την ομαδοποίηση και την ανάλυση μεγάλων συνόλων δεδομένων. Ο συνδυασμός της *NER* και του *BERTopic*, μιας τεχνικής μοντελοποίησης θεμάτων βασισμένης στο *BERT*, έχει αποδειχθεί αποτελεσματικός στον εντοπισμό και την κατηγοριοποίηση εργασιών και θεμάτων από δεδομένα κοινωνικών μέσων, όπως αποδείχθηκε σε πρόσφατες μελέτες που ανέλυαν *tweets* σχετικά με το *ChatGPT*.

Καθώς, οι *LLMs* συνεχίζουν να εξελίσσονται, η ικανότητά τους να εξάγουν με ακρίβεια πληροφορίες θα βελτιωθεί. Τα μελλοντικά μοντέλα αναμένεται να έχουν καλύτερη κατανόηση του πλαισίου, μειώνοντας την ασάφεια και βελτιώνοντας την ακρίβεια των εξαγόμενων

πληροφοριών. Αυτό θα είναι ιδιαίτερα επωφελές σε τομείς που απαιτούν υψηλή ακρίβεια, όπως η ανάλυση νομικών εγγράφων και η επιστημονική έρευνα. Η ικανότητα των *LLMs* να επεξεργάζονται και να αναλύουν πληροφορίες σε πραγματικό χρόνο θα ανοίξει νέες δυνατότητες για δυναμικές και διαδραστικές εφαρμογές. Για παράδειγμα, η εξόρυξη πληροφοριών σε πραγματικό χρόνο θα μπορούσε να βελτιώσει την εξυπηρέτηση πελατών παρέχοντας άμεσες, σχετικές με το πλαίσιο απαντήσεις σε ερωτήματα χρηστών. Ομοίως, στη διαχείριση έκτακτων αναγκών, οι *LLMs* θα μπορούσαν να αναλύουν ζωντανές ροές δεδομένων για την παροχή ενημερωμένων πληροφοριών και συστάσεων. Η ενσωμάτωση των *LLMs* με άλλες αναδυόμενες τεχνολογίες, όπως η όραση υπολογιστών και το Διαδίκτυο των πραγμάτων (*IoT*), θα διευρύνει περαιτέρω το πεδίο εφαρμογής τους. Για παράδειγμα, ο συνδυασμός των *LLMs* με οπτικά δεδομένα από συσκευές *IoT* θα μπορούσε να επιτρέψει ολοκληρωμένα συστήματα παρακολούθησης σε έξυπνες πόλεις, ικανά να εξάγουν και να αναλύουν πληροφορίες από πολλαπλές πηγές ταυτόχρονα. Ένας από τους σημερινούς περιορισμούς των *LLMs* είναι η φύση τους ως "μαύρο κουτί", η οποία καθιστά δύσκολο να κατανοηθεί πώς καταλήγουν σε συγκεκριμένες εξόδους. Η μελλοντική έρευνα είναι πιθανό να επικεντρωθεί στη βελτίωση της επεξηγηματικότητας και της διαφάνειας αυτών των μοντέλων, η οποία θα είναι ζωτικής σημασίας για την υιοθέτησή τους σε ευαίσθητους τομείς όπως η υγειονομική περίθαλψη και η χρηματοδότηση. Τεχνικές από τον τομέα της εξηγήσιμης τεχνητής νοημοσύνης (*Explainable AI - XAI*) θα διαδραματίσουν σημαντικό ρόλο στο πλαίσιο αυτό. Καθώς τα μοντέλα *LLMs* γίνονται πιο διαδεδομένα, η διασφάλιση της ηθικής και υπεύθυνης χρήσης τους θα είναι ζωτικής σημασίας. Αυτό περιλαμβάνει την αντιμετώπιση ζητημάτων που σχετίζονται με την προστασία της ιδιωτικής ζωής των δεδομένων, την προκατάληψη και την πιθανή κατάχρηση των πληροφοριών.

Το μέλλον των *LLMs* στην εξαγωγή πληροφοριών είναι πολλά υποσχόμενο, με σημαντικές εξελίξεις που αναμένονται στην ακρίβεια, τις δυνατότητες πραγματικού χρόνου, την ενσωμάτωση με άλλες τεχνολογίες και την επεξηγηματικότητα. Ωστόσο, η αντιμετώπιση των ηθικών ανησυχιών και η διασφάλιση της υπεύθυνης χρήσης θα είναι υψίστης σημασίας. Καθώς αυτά τα μοντέλα συνεχίζουν να εξελίσσονται, είναι έτοιμα να φέρουν επανάσταση στον τρόπο με τον οποίο εξάγουμε και χρησιμοποιούμε πληροφορίες σε διάφορους τομείς, προωθώντας την καινοτομία και βελτιώνοντας την αποτελεσματικότητα σε πολυάριθμες εφαρμογές. (Chiarello, 2024)

Συμπεράσματα

Η έρευνα που παρουσιάζεται στη διπλωματική εργασία επικεντρώνεται στην ανάλυση της αγοράς εργασίας μέσω των διαδικτυακών αγγελιών, χρησιμοποιώντας προηγμένες τεχνολογίες εξαγωγής και επεξεργασίας πληροφοριών. Τα βασικά συμπεράσματα που εξάγονται από την παρούσα διπλωματική είναι ότι οι τεχνολογίες μηχανικής μάθησης και τεχνητής νοημοσύνης έχουν αποδειχθεί εξαιρετικά αποτελεσματικές στην αυτόματη ανάλυση και κατηγοριοποίηση δεδομένων από διαδικτυακές αγγελίες εργασίας. Τα συστήματα που αναπτύχθηκαν κατάφεραν να κατηγοριοποιήσουν και να αναλύσουν τις αγγελίες με υψηλή ακρίβεια, παρέχοντας πολύτιμες πληροφορίες για τις απαιτούμενες δεξιότητες και τις διαρθρωτικές ανάγκες της αγοράς εργασίας. Η χρήση των *ML* και *AI* μοντέλων επέτρεψε την αντιμετώπιση της πολυπλοκότητας και της ποικιλομορφίας των δεδομένων των αγγελιών εργασίας, προσφέροντας μια αποδοτική λύση για την εξαγωγή χρήσιμων πληροφοριών από μεγάλα και ετερογενή σύνολα δεδομένων. Ένας σημαντικός περιορισμός της έρευνας είναι η φύση των *ML* μοντέλων ως «μαύρα κουτιά», που καθιστά δύσκολη την κατανόηση του τρόπου λήψης των αποφάσεων. Επιπλέον, η έλλειψη τυποποίησης και η παρουσία θορύβου στα δεδομένα των αγγελιών αποτελούν προκλήσεις για την ακριβή ανάλυση.

Για την βελτίωση διαφάνειας και κατανόησης των αποτελεσμάτων των μοντέλων, προτείνεται η περαιτέρω έρευνα και ανάπτυξη τεχνολογιών *Explainable AI (XAI)*. Αυτό θα επιτρέψει στους χρήστες να κατανοούν καλύτερα τις διαδικασίες λήψης αποφάσεων των συστημάτων *AI* και να εμπιστεύονται περισσότερο τα αποτελέσματα. Οι αναπτυγμένες τεχνολογίες μπορούν να χρησιμοποιηθούν από επιχειρήσεις και οργανισμούς για την καλύτερη κατανόηση των αναγκών της αγοράς εργασίας και τη λήψη πιο ενημερωμένων αποφάσεων προσλήψεων. Οι εταιρείες μπορούν να αξιοποιήσουν τα συστήματα αυτά για να εντοπίζουν γρήγορα τις απαραίτητες δεξιότητες και να προσαρμόζουν τις στρατηγικές προσλήψεων τους αναλόγως. Επιπροσθέτως, προτείνεται η εφαρμογή αλγορίθμων μηχανικής μάθησης, όπως το συνεργατικό φιλτράρισμα και το φιλτράρισμα βάσει περιεχομένου, για την παροχή εξατομικευμένων προτάσεων εργασίας στους υποψηφίους. Αυτό θα συμβάλλει στην αύξηση της αποτελεσματικότητας της διαδικασίας αναζήτησης εργασίας και θα βελτιώσει την εμπειρία των υποψηφίων. Τέλος, συνιστάται η συνεχιζόμενη έρευνα στις μελλοντικές δυνατότητες των μεγάλων γλωσσικών μοντέλων (*LLMs*) στην εξαγωγή πληροφοριών, καθώς και η διερεύνηση της ηθικής και της υπεύθυνης χρήσης των τεχνολογιών αυτών για να εξασφαλιστεί η αποδοχή και η αποτελεσματικότητα τους σε ποικίλες εφαρμογές.

Η παρούσα διπλωματική εργασία αποδεικνύει τη σημαντικότητα της συνδυαστικής χρήσης των τεχνολογιών *ML*, *AI*, και *NLP* στην ανάλυση της αγοράς εργασίας μέσω διαδικτυακών αγγελιών, προτείνοντας παράλληλα μεθόδους για την περαιτέρω βελτίωση και εφαρμογή αυτών των τεχνολογιών.

ΒΙΒΛΙΟΓΡΑΦΙΑ

A. Αγγλική

Arif Ali Mughal,(2018). Artificial Intelligence in Information Security: Exploring the Advantages, Challenges, and Future Directions

Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python. O'Reilly Media, Inc.

Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation*, 133, 103002. Available online: Elsevier.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation?. *Technological forecasting and social change*, 114, 254-280.

Friedl, J. E. F. (2006). Mastering regular expressions. O'Reilly Media, Inc.

Géron, Aurélien. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., 2017.

Ghani, R. A., Hamid, N. H. A., Abdullah, N. A. B., & Ibrahim, Z. (2017). Web scraping: A survey. *Computers & Security*, 68, 145-157.

Hjorth, K., & Zhang, Y. (2016). Web scraping. *ACM Computing Surveys (CSUR)*, 49(1), 1-33.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R*. Springer Science & Business Media.

Jurafsky, D., & Martin, J. H. (2021). *Speech and language processing (3rd ed.)*. Pearson.

Kowsher, Md., Islam Sanjid, Md. Z., Das, A., Ahmed, M., & Hossain Sarker, Md. M. (2020). Machine Learning and Deep Learning based Information Extraction from Bangla Names. *Procedia Computer Science*, 178, 224–233. <https://doi.org/10.1016/j.procs.2020.11.024>

Krotov, V., Johnson, L., & Silva, L. (2020). Legality and Ethics of Web Scraping. *Communications of the Association for Information Systems*, 47, 539–563.

Mehra, A., Misra, A., & Demirkan, H. (2020). A review of job posting analysis using artificial intelligence. *Artificial Intelligence Review*, 54(2), 1143–1180

Nadeau, D., & Sekine, S. (2007). A survey of named entity recognition and classification. *Linguistics and Language Resources*

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson Education Limited.

Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). Using Text Mining Techniques for Extracting Information from Research Articles. *Studies in Computational Intelligence*, 373–397. https://doi.org/10.1007/978-3-319-67056-0_18

Zhi Hong, Logan Ward ,Kyle Chard, Ben Blaiszik ,Ian Foster(2021). Challenges and Methodologies for Web Information Extraction: A Review

B. Ιστοσελίδες

Astera Software, Data Extraction Tools: Here's Everything You Need To Know στο: <https://www.youtube.com/watch?v=nJLUYIAUqIQ> (προσπελάστηκε στις 26/09/2023)

Coursera ,What is Natural Language Processing? Definition and Examples διαθέσιμο στο URL: <https://www.coursera.org/articles/natural-language-processing>(προσπελάστηκε στις 19/03/2024)

CSU GLOBAL,Why is AI important? στο: <https://csuglobal.edu/blog/why-ai-important> (προσπελάστηκε στις 5/07/2021)

Datarade, Best Job Search API's to use in 2023,στο: <https://datarade.ai/top-lists/best-job-search-apis> (προσπελάστηκε στις 26/06/2023)

Egnati Simply Intelligent, Information extraction, στο: <https://www.engati.com/glossary/information-extraction> (προσπελάστηκε το 2021)

GlassAPI,(2008-2024), στο: <https://www.glassdoor.com/developer/companiesApiActions.htm>

Google Cloud ,Document AI στο : <https://cloud.google.com/document-ai>

Hevo Data. , The Modern Data Integration Platform στο: <https://hevodata.com/>

IBM, What is machine learning (ML)?, στο : <https://www.ibm.com/topics/machine-learning>

Indeed, στο : https://www.indeed.com/jobs?q=Developer&from=mobRdr&utm_source=%2Fm%2F&utm_medium=redir&utm_campaign=dt&vjk=8ce86cc81a001a76

InDevelopers στο : <https://developer.linkedin.com/> (προσπελάστηκε το 2024)

Integrate.io., Top 10 Data Extraction Tools for 2024 στο <https://www.integrate.io/>
(προσπελάστηκε στις 9/02/2024)

Linkedin, στο: <https://www.linkedin.com/advice/1/how-can-you-use-web-scraping-extract-data-from-job> (προσπελάστηκε το 2024)

Machine Learning Mastery, στο : <https://machinelearningmastery.com/> (προσπελάστηκε το 2024)

Microblink., 10 Best Data Extraction Software and Tools for 2024, στο: <https://microblink.com/>. (προσπελάστηκε στις 10/01/2024)

Parsehub, What is Web Scrapping and what is it Used For? στο URL: <https://www.parsehub.com/blog/what-is-web-scraping/> (προσπελάστηκε στις 14/04/2023)

Rivery,14 Best Data Extraction Tools to Choose in 2024 στο: <https://rivery.io/>
(προσπελάστηκε 11/01/2024)

Simplilearn, Advantages and Disadvantages of Artificial Intelligence [AI] στο:<https://www.simplilearn.com/advantages-and-disadvantages-of-artificial-intelligence-article>,(προσπελάστηκε 15/03/2024)

Society for Human Resource Management, Challenges of Big Data Analytics in HR

TechTarget, EnterPrise AI, DEFINITION natural language processing(NPL)διαθέσιμό στο link: <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML>

TechTarget, EnterPrise AI, What is machine learning and how does it work? In- depth guide στο : <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML>(προσπελάστηκε το 2024)

TrustRadius, List of Top Data Extraction Tools 2024, στο: <https://www.trustradius.com/> (προσπελάστηκε στις 14/02/2024)

UiPath, Robotic Process Automation στο: <https://www.uipath.com/>

upGrad, Top Advantages and Disadvantages of Machine Learning in 2024 στο : <https://www.upgrad.com/blog/advantages-and-disadvantages-of-machine-learning/>,(προσπελάστηκε στις 24/08/2023)

WGU,All the Benefits Of Artificial Intelligence στο : <https://www.wgu.edu/blog/benefits-artificial-intelligence2204.html> ,(προσπελάστηκε στις 25/04/2022)

Wikipedia, RSS διαθέσιμο στο link: <https://el.wikipedia.org/wiki/RSS> (προσπελάστηκε στις 22/05/2023)

Wikipedia, Η ελεύθερη εγκυκλοπαίδεια, Εξόρυξη πληροφορίας, στο: https://el.wikipedia.org/wiki/%CE%95%CE%BE%CE%B1%CE%B3%CF%89%CE%B3%CE%AE_%CF%80%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%B9%CF%8E%CE%BD (προσπελάστηκε στις 31-10-2024)