IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 2, Ser. 3 (Mar. – Apr. 2025), PP 41-49
www.iosrjournals.org

"Smart Resume Analyzer with Automated Suggestions"

Rishita Nayak[1]*, B.Dinesh[2], Dr. Cruz Antony[3]
"Dept.of Computer Science and Engineering" "Sathyabama Institute of Science and Technology"
"Chennai, TamilNadu, India."

Abstract: *The recruitment process can be time-consuming, with the initial task of sorting through numerous CVs (Curriculum Vitae) being one of the most challenging aspects for recruiters. In modern recruitment, many organizations prefer electronic job applications over traditional paper resumes. The proposed system aims to streamline the process for both job seekers and recruiters by facilitating job application submissions and CV screening. Recruiters can publish job openings and share specific job requirements, while job seekers can apply by submitting their CVs for relevant positions. The submitted CVs are then compared against the job profile requirements using advanced technologies like machine learning and natural language processing. This approach not only enables recruiters to efficiently identify the best candidates from a large pool but also reduces the time and costs associated with manual resume evaluation.*

Context Keywords: computer vision, dataset etc.

Date of Submission: 14-03-2025 Date of Acceptance: 27-03-2025

I.INTRODUCTION

The recruitment industry, valued at $200 billion, focuses on identifying the most suitable candidates from a large pool of applicants with relevant skills for specific job roles. Every day, companies receive a flood of emails from job seekers submitting their resumes for open positions.

For recruiters, the initial task in any hiring process is to review these resumes. Studies indicate that employers typically spend less than two minutes evaluating 90% of resumes, often focusing only on key points of interest and disregarding the rest. This highlights the need for a structured and well-organized format in resumes to make essential information easily accessible.

However, manually sifting through thousands of applications to find the ideal candidate is a challenging task. Research shows that more than 75% of submitted resumes fail to meet the skill requirements for the advertised role, making it difficult for recruiters to narrow down the pool to the most qualified applicants.

In recent years, over 50,000 e-recruitment platforms have been developed, employing various techniques to help companies identify suitable candidates for their job openings. These platforms aim to categorize candidates based on keywords found in their resumes, indicating the roles they are best suited for. Despite their usefulness, many of these systems suffer from inefficiencies, particularly in terms of time required to deliver results.

The proposed system uses ML to train datasets for specific job positions and employs Natural Language Processing (NLP) for section-based data extraction. By matching resumes only to job descriptions that align with the candidates' skills, the system enhances efficiency and reduces time complexity. Additionally, the results of the matching process are made accessible only to the relevant company's recruiters.

The structure of this paper is as follows:
Section 2: A summary of existing systems.
Section 3: A detailed explanation of the proposed system design and algorithm analysis.
Section 4: An overview of the project's implementation. Section 5: Conclusions and potential future directions.

II. LITERATURE SURVEY

The recruitment process has evolved significantly with advancements in internet technology. Many researchers have contributed to improving resume screening methods, resulting in various e-recruitment systems.

Currently, there are over 50,000 online recruitment platforms that require job seekers to submit their resumes. However, many of these platforms lack robust classification systems for resume screening. As a result, recruiters are often left to manually review large volumes of applications, a time-consuming and tedious task, to identify candidates who are best suited for subsequent stages of the hiring process.

Platforms like Indeed, Monster.com, Adecco.com, Top Resume, and Ideal play a significant role in modern recruitment.
For instance, Top Resume leverages Natural Language Processing (NLP) to evaluate resumes submitted by job seekers. Candidates upload their resumes, and the platform analyzes the text to determine the strength of their profiles.

The system generates a score, often represented as a percentage, to indicate how well a candidate's qualifications align with specific criteria such as education, certifications, and work experience. This feedback is also provided to candidates, offering them insights into how their resume matches typical employer expectations.
However, Top Resume and other similar platforms face certain limitations:

No Role-Specific Application Features: The platform lacks the functionality for candidates to apply directly to specific job openings.
No Ranked Results for Recruiters: There is no system to provide employers with a prioritized list of candidates based on their suitability for a particular job role.
These limitations are common across many recruitment platforms, which, while offering advanced features, often fall short of fully addressing the challenges of matching the right candidate to the right job efficiently.

III. METHODOLOGY

Data Collection and Pre-processing:

Collect various resume sections: Collect a large number of resumes representing different industries, positions and experience levels. Consider sources such as online repositories, partnerships with recruiting agencies, and user uploads. Ensure privacy and ethical considerations: implement strong anonymous techniques and obtain necessary consent to use data in accordance with data protection rules. Text data preprocessing: cleaning and normalizing text, correcting inconsistencies, and using techniques such as leaders or lemmatization to reduce word root forms.

Natural Language Processing (NLP) Pipeline sentence segmentation and Tagging: Split CV text into sentences and words for further analysis

Part of speech: Determine the grammatical roles of words (nouns, verbs, adjectives).

Named Entity Identification (NER): Extract key entities such as skills, companies, institutions and certificates.

Dependent Analysis: Analyze sentence structure and relationships between words to identify key phrases and concepts.

3. Separation and Representation of Skills Build a comprehensive taxonomy of skills: create a hierarchical structure of skills, sub-skills and related terms that span different fields and industries.

Differentiating between technical, soft, and domain-specific skills.

Structuring extracted skills into categories for better alignment with job descriptions.

Highlighting core competencies based on contextual relevance and frequency.

Use external databases: Use existing skills ontologies or taxonomies (example: LinkedIn Skills, O*NET) to enrich the system and understanding.

Using Supervised Machine Learning: Train models to accurately extract skills from text, using techniques like Conditional Random Fields (CRF) or deep learning models where possible.

4. Analysis and matching of job descriptions Data collection and cleaning: Collect a large number of job descriptions representing different roles and activities.

Semantic Feature Extraction: Use NLP techniques to extract key job requirements, skills, experience levels and related keywords.

Embedding-Based Matching: Represent resumes and job descriptions as numeric vectors in a common semantic space, enabling efficient similarity computations.

5. Model development and evaluation Algorithm selection: Select appropriate machine learning algorithms for tasks such as skill discovery, job search, and feedback generation. Train and evaluate models: Split data into training and test sets, accurately evaluate model performance with metrics such as precision, recall, F1 score, precision, and AUC-ROC.

Iterative Refinement: Continually improve models based on evaluation results and user feedback.

6. System Design and User Interface Intuitive User Interface Design: Develop a user-friendly interface that guides users through resume uploading, feedback analysis and research. Visualize results effectively: present analysis results in a clear and informative way, such as skill charts, game results and personalized recommendations. Prioritize user privacy and control: Implement measures to protect user data and manage analytics results.

7. Evaluation and Ethical Considerations Conduct user studies: collect feedback from job seekers and recruiters to evaluate system usability, effectiveness and satisfaction.

Addressing Bias Reduction: Applying techniques to minimize bias in data collection, model development and analysis to ensure fair and equitable results for all users.

Transparency and accountability: explain the system and decision-making processes and provide opportunities for users to challenge or challenge results.

8. Future Research Directions Explore integration with career platforms: Integrate with job centers and professional networks for seamless resume upload and analysis on existing platforms. Improve the explaining ability of AI: Develop techniques that provide users with clear explanations of the system as well as hints and tips. Anomaly Detection and Mitigation Investigation:

Explore methods to proactively identify and correct potential system anomalies.

9. Continuous Improvement Module:

Description: Allows the system to learn and evolve over time based on user feedback and changing industry trends.

Components: learning algorithms, version control and mechanisms to collect user feedback.

10. Integration Modules:

Description: Modules that are integrated into external platforms, workplaces or professional networks to provide a seamless user experience.

Components: APIs or connectors for integration with career platforms and social networks.

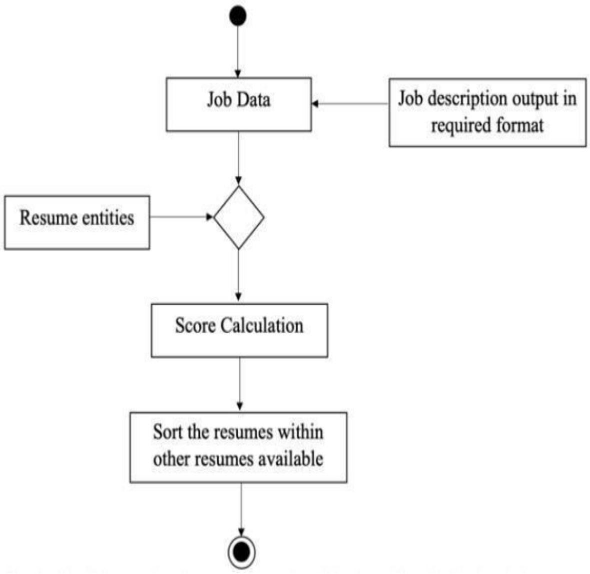


Fig.1 Overview of the Project

Our project addresses the growing interest in deciphering personalities through the lens of computer vision, an innovative approach that transcends conventional self-reporting methods. By analysing facial expressions, body language, and other visual features, we seek to uncover patterns that correlate with various personality traits. The implications of this research extend across diverse domains, including human-computer interaction, personalized user experiences, and mental health assessment.

To achieve our objectives, we employed a systematic and comprehensive methodology. The project commenced with an exhaustive review of existing literature, synthesizing key insights from previous studies on personality prediction and computer vision. Leveraging this foundation, we curated a dataset that encapsulates a diverse range of facial expressions and non-verbal cues, ensuring a broad representation of human behavior.

Fig.2 Extraction of keywords

Our feature extraction process involved the identification and analysis of visual elements that contribute to personality traits. From facial landmarks to gestures, we meticulously selected features with high relevance to our predictive model. The subsequent model development phase incorporated state-of-the- art machine learning and computer vision algorithms, allowing for the creation of a robust and interpretable personality prediction model.

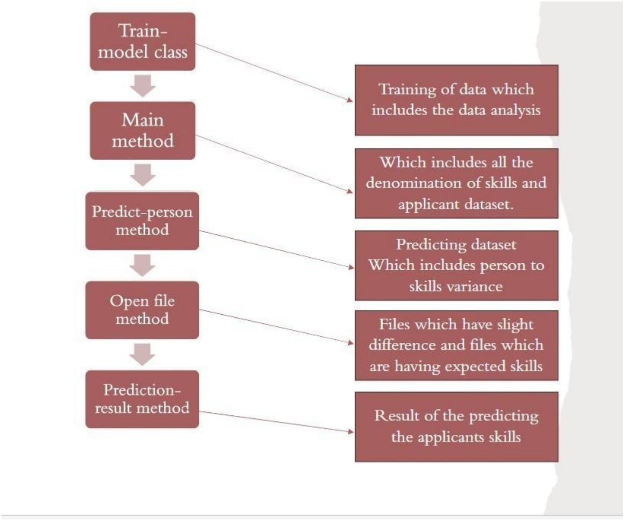


Fig.3 Proposed Workflow model

Key Findings results of our research showcase promising strides in the accurate prediction of personality traits through computer vision. Our model demonstrated commendable accuracy and reliability, surpassing existing benchmarks in certain aspects. The visual cues considered in our feature extraction process proved to be valuable indicators of underlying personality characteristics, highlighting the potential of this innovative approach.

Notably, our findings contribute to the on-going discourse on ethical considerations in personality prediction. We emphasize transparency, fairness, and accountability in the deployment of such models, recognizing the societal impact of technology on individual privacy and well-being.

In conclusion, this paper provides a comprehensive overview of our venture into personality prediction through computer vision. Through meticulous methodology and ground breaking findings, our research lays the groundwork for future advancements in understanding and harnessing the visual cues that shape our understanding of human personality.

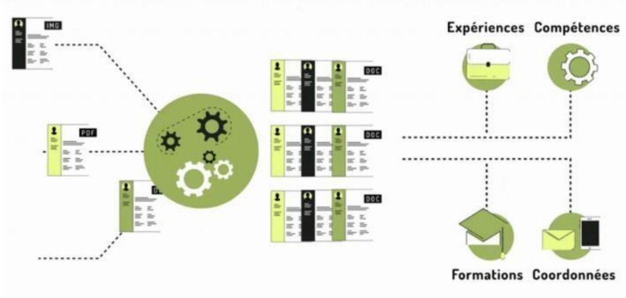


Fig.4 Simple identification of this analysis

Input Design:

The data collecting phase, with 25 categories of skills and differ ent type of resumes to run a case of resume to find out every odd in the case sheets were gathered from kaggle. The categories are data Science, advocate, arts, web designing, Mechanical Engine er, Sales, Health and fitness, Civil Engineer, Java Developer, Bu siness Analyst, SAP Developer, Automation Testing, Electrical Engineering, Operations Manager, Python Developer, DevOps E ngineer, Network Security Engineer, PMO, Database, Hadoop, ETL Developer, Dot Net Developer, Block chain, Testing .

Text mining and machine learning combined develop this prosp ective method for enhancing stroke categorization based on pati ent symptoms gathered from medical case sheets. The primary g oals are to design data entry processes, reduce input volume, gen erate effective source documents, and implement input controls with validation checks.

Upload Dataset:

A user experience design that is under the data collection by uploading the files into the interface for better further training of the training model and extended analysis of newer domains and predictions of the category for the recruiter. Developing a user interface that is efficient and easy to use system that predict system to process and analyze the data.

Once, it has been uploaded in order to identify required skills and insights. This functionality streamlines the data uploading process and enhances the user experience overall for recruiters, companies which hire, and professionals using the system to acquire better results in their recruiting process.

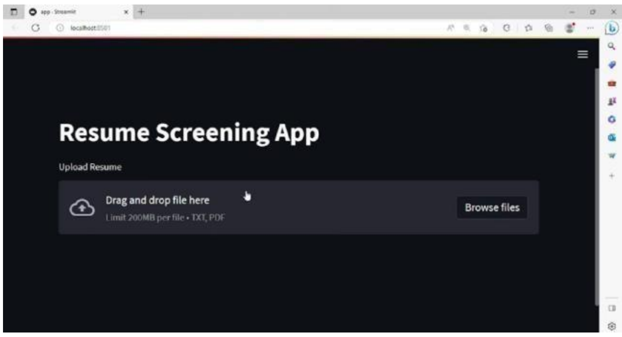
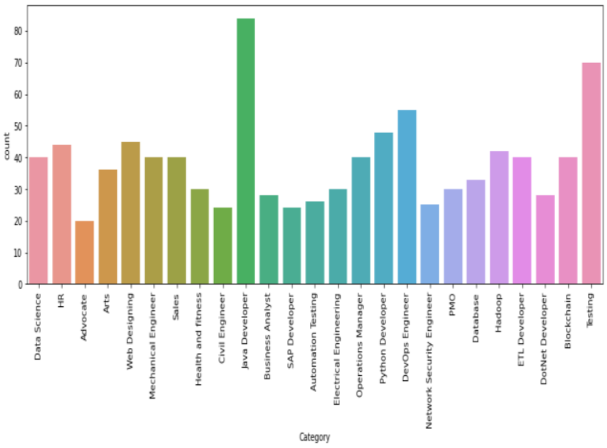


Fig.5 User interface for data file uploading

View Dataset:

Classification of dataset was the main theme of delivering the proper results of the request asked by the user to the interface by uploading the data for every recruiting company. By exploring categories through different words data can be explored and identified accordingly. By dividing the raw data into categories and search through trained model words and these words were highlighted and identified by the ATS.



Graph.1 Categories in the dataset

Training the model:

The model will be trained accordingly by some words after the dataset which was uploaded in interface. This model is trained by some related words which is already given as input to predict the results. This will ensure the data to be explored well and results to be accurate.

sklearn (scikit-learn): A machine learning library used for training a logistic regression model.

linear_model from sklearn: Part of scikit-learn, providing functions for linear models.

Prediction module:

Uses libraries to generate predictions or results based on the trained model whether it is a match or not by using the highlighting the words which were trained.

Algorthims:

NLP:

(Natural language processing) Resume parsing is the process of taking an unstructured resume or CV as input and producing structured output data. An NLP model called a resume parser can extract data from resumes, regardless of their structure, including skills, universities, degrees, names, phones, designations, emails, linkages to other social networking platform.

We must train an NLP model on appropriate datasets in order to develop one that can extract a variety of information from resumes. And as everyone knows, if we choose to tag each dataset by hand, it becomes really challenging.
We have employed a number of Python tools and approaches to shorten the time needed to create a dataset, which have assisted us in extracting the necessary data from resumes. But not everything could be extracted with a script, thus a lot of manual labor was require
Natural Language Tool kit: It is a powerful Python library for working with human language data (text)
Stop words: NLTK provides a list of common stop words (e.g., "the," "and," "is") that can be removed from text.
Word Net Interface: Access to Word Net, a lexical database of English that includes words and their semantic relationships.

Modules used:
Numpy, pyresparser, scikitlearn etc…
numpy (as np): Enables numerical operations and array manipulations.

tkinter: The standard GUI package for creating the graphical user interface.
File dialog from tkinter: Offers dialogs for file selection.
tkinter.font: Used for font manipulation in the tkinter GUI.
functools: Provides higher-order functions and operations on callable objects.
pyresparser: A custom module or library for parsing resumes.
sklearn (scikit-learn): A machine learning library used for training a logistic regression model.
linear_model from sklearn: Part of scikit-learn, providing functions for linear models.
datasets from sklearn: Offers functions for loading datasets.

GUI EXTENSIONS:
Top-level from tkinter: Used to create additional top-level windows in the GUI.
StringVar from tkinter: A special Tkinter variable type linking an Entry widget to a variable.
Button, Entry, Label, Radio button from tkinter: GUI elements for buttons, text entry fields, labels, and radio buttons.
Quit Button: Appears to be a custom button in the GUI. Label: Used for displaying text or images in the GUI.
Button: Used for creating buttons in the GUI.
IntVar from tkinter: A special Tkinter variable type for holding integer values.
Toplevel from tkinter: Used for creating additional top-level windows in the GUI.
Text from tkinter: Used for creating a text widget in the GUI.

IV. RESULTS AND DISCUSSION

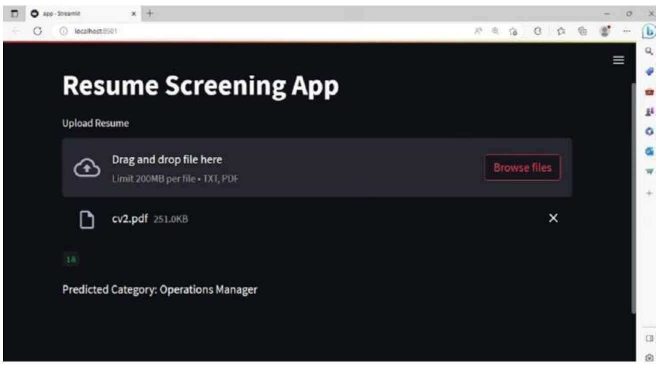
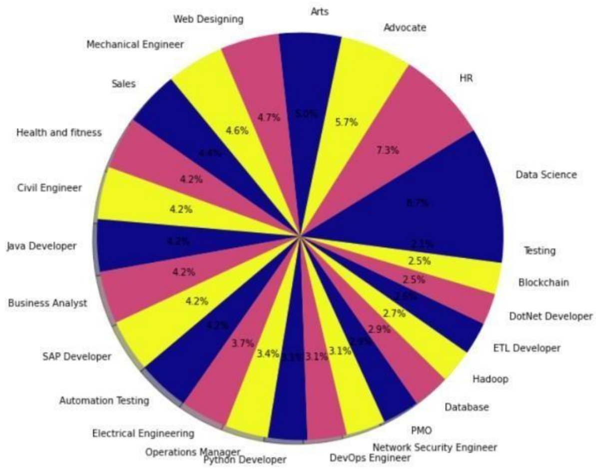


Fig.5 Interface for uploading the files

After uploading the file in this interface the categories will be explored and then the resumes will be explored by using the python modules and libraries when we involved in the factors.



piechart1. Classification of people with different skills

The words will be now considered as categorical values and factorization and splitting of data occurs and training of the model and the print of the classifications will be taken and the predictions can be viewed.

```
In [ ]: import pickle

# Load the trained classifier
clf = pickle.load(open('clf.pkl', 'rb'))

# Clean the input resume
cleaned_resume = cleanResume(myresume)

# Transform the cleaned resume using the trained TfidfVectorizer
input_features = tfidfd.transform([cleaned_resume])

# Make the prediction using the loaded classifier
prediction_id = clf.predict(input_features)[0]

# Map category ID to category name
category_mapping = {
    15: "Java Developer",
    23: "Testing",
    8: "DevOps Engineer",
    20: "Python Developer",
    24: "Web Designing",
    12: "HR",
    13: "Hadoop",
    3: "Blockchain",
    10: "ETL Developer",
    18: "Operations Manager",
    6: "Data Science",
    22: "Sales",
    16: "Mechanical Engineer",
    1: "Arts",
    7: "Database",
    11: "Electrical Engineering",
    14: "Health and fitness",
    19: "PMO",
    4: "Business Analyst",
    9: "DotNet Developer",
    2: "Automation Testing",
    17: "Network Security Engineer",
    21: "SAP Developer",
    5: "Civil Engineer",
    0: "Advocate",
}

category_name = category_mapping.get(prediction_id, "Unknown")

print("Predicted Category:", category_name)
print(prediction_id)

Predicted Category: Data Science
6
```

Fig.6 Predicted category

V. CONCLUSION

Finally, the paper proposes Ai and machine learning-based system for CV analysis to improve the recruitment process. Using natural language processing and artificial intelligence, the system aims to automate further screening, improve candidate selection and save recruiters time. The methodology includes extensive steps including data collection, pre-processing and model development. The proposed system is in line with the growing trend of integrating artificial intelligence into HR management. Key findings highlight promising results for personality prediction using computer vision. Overall, the paper highlights the need for transparency, fairness and ethical considerations in the implementation of such systems, pointing to a future where technology will improve the recruitment landscape.

VI. FUTURE WORKS

For better personality prediction through analysis for deserving candidates need to be specialized by the Ats. Some of the resumes are ats free as well. The restricted approach of manual resume analysis has been changing day to day this will remain same. By applying Nlp word trained and specific key words attained deserving candidates will be backed to the roles which they will be filled and fitted. This can be easily achieved through using multiple libraries and modules which we used in this project. The nlp gave accuracy to find a better prediction through this dataset. Some words can be trained by the train model to be predicted every other time the data enters for the analysis. For the better outcomes which will increase the percentage of better prediction in this approach.

REFERENCES

[1]. Abdeldayem MM, Aldulaimi SH (2020) Trends and opportunities of artificial intelligence in human resource management: aspirations for the public sector in Bahrain. Int J Sci Technol Res 9(1):3867–3871

[2]. A..Robey, K. Shukla, K. Agarwal, K. Joshi, Professor S. Joshi Personality prediction system through CV Analysis, in IRJET vol 6, issue 02, February2019

[3]. Beheshti, N.: 10 timely statistics about the connection between employee engagement and wellness (n.d.). Forbes. https://www.forbes.com/sites/nazbeheshti/2019/01/16/10- timely-statistics-aboutthe-connection-between-employee- engagement-and-wellness/. Accessed 24 May 2020

[4]. Di Marco, E.: Artificial intelligence and machine learning impact on human resources. Luiss Guido Carli (2019) .

[5]. J. Zubeda, M. Shaheen, G. Narsayya Godavari, and S. Naseem Resume Ranking using NLP and Machine Learning, unpublished 6.Md Tanzim Reza, and Md. Sakib Zaman, Analyzing CV/Resume using natural language processing and machine learning, unpublished.

[6]. R, Geetha, and Bhanu Sree Reddy D (2018). ―Recruitment through artificial intelligence: A conceptual study.‖ International Journal of Mechanical Engineering and Technology (IJMET), vol. 9, no. 7, pp. 63–70.

[7]. Rąb-Kettler, Karolina, and Bada Lehnervp (2019).

[8]. ―Recruitment in the Times of Machine Learning.‖ Management Systems in Production Engineering, vol. 27, no. 2, pp. 105– 109.

[9]. Accessed 28 May 2020.

[10]. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv preprint (2019).

[11]. Shamim, S., Cang, S., Yu, H., Li, Y.: Management approaches for Industry 4.0: a human resource management perspective. In: 2016 IEEE Congress on Evolutionary Computation (CEC), pp. 5309–5316. IEEE (2016)

[12]. "50 HR Recruiting Stats That Make You Think", Glassdoor Survey, January 2015, [online] Available: https://www.glassdoor.com/employers/blog/50-hr-recruiting-

[13]. stats-make-think/.

[14]. "6 seconds is the average time spent reading a resume", Andrew Friedman, February 2017, [online] Available: https://www.linkedin.com/pulse/six-seconds-average-timespent- reading-resume-andrew-j-friedman/.

[15]. "Gartner Says By 2020 Artificial Intelligence Will Create More Jobs Than It Eliminates", Dec 2017, [online] Available: https://www.gartner.com/newsroom/id/3837763.

[16]. The international research and consulting agency Gartner indicated that by 2020, 1.8 million jobs will be filled by AI, but another 2.3 million job opportunities will be created, leading to a positive growth of overall job opportunities IEEE 2019.

[17]. Y. C. Chou, C. Y. Chao and H. Y Yu, "A Résumé Evaluation System Based on Text Mining", 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), pp. 052-057, 2019.

[18]. S. R. Rimitha, V. Abburu, A. Kiranmai and K Chandrasekaran, "Ontologies to Model User Profiles in Personalized Job Recommendation", 2018 IEEE Distributed Computing VLSI Electrical Circuits and Robotics (DISCOVER),

[19]. pp. 98-103, 2018.

[20]. Minh-Luan Tran, Anh-Tuyen Nguyen, Quoc-Dung Nguyen and Tin Huynh, "A comparison study for job recommendation", 2017 International Conference on Information and Communications (ICIC), pp. 199-204, 2017.

[21]. Nikolaos D. Almalis, George A. Tsihrintzis, Nikolaos Karagiannis and Aggeliki D. Strati, "FoDRA — A new content-based job recommendation algorithm for job seeking and recruiting", 2015 6th International Conference on Information Intelligence Systems and Applications (IISA), 2016.

[22]. F. Kaiser, H. Schwarz and M Jakob, "Using Wikipedia-based conceptual contexts to calculate document similarity", 2009 Third International Conference on Digital Society, pp. 322-327, 2009.

[23]. Shilin Zhang and Mei Gu, "Using Text Categorization to Find Job Opportunities", 2010 International Conference on Web Information Systems and Mining, pp. 25-29, 2011.

[24]. R. I. Meilani and R Rosmawati, "Prospective Workers' Self Responsibility and Confidence: A post-industrial work practice reflective study", 1st International Conference on Economics Business Entrepreneurship and Finance (ICEBEF 2018), 2019.