“AI – Enabled HR Analytics”

Kahan Jash
Malhar Shinde
Parth Parikh

Guided by Komal Lamba
May 2024

INDEX

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Prologue

In the ever-evolving landscape of human resources management, organizations are constantly seeking innovative solutions to streamline processes, enhance efficiency, and make data-driven decisions. Enter AI-enabled HR Analytics – a groundbreaking application poised to revolutionize traditional HR procedures. This cutting-edge technology harnesses the power of artificial intelligence, natural language processing, and data analytics to automate and optimize various aspects of the HR lifecycle.

AI-enabled HR Analytics offers a comprehensive suite of features designed to empower HR professionals with actionable insights and facilitate strategic decision-making. One of its key functionalities is resume screening with AI precision. By leveraging advanced NLP algorithms, the application ensures a thorough and impartial evaluation of candidate resumes, expediting the selection process while minimizing bias.

Furthermore, the integration of an AI-driven chatbot redefines candidate engagement, providing a seamless and personalized experience for applicants. This not only accelerates administrative tasks such as interview scheduling but also enhances interaction throughout the recruitment process.

A unique feature of AI-enabled HR Analytics is its capability for sentiment analysis through multimedia assessment. By analyzing images and videos, the application offers a deeper understanding of candidate sentiment, enabling HR managers to make more informed hiring decisions.

Moreover, predictive analytics empowers organizations to anticipate and address potential retention issues by analyzing historical data. Armed with insights into employee behavior and trends, HR specialists can proactively implement strategies to improve employee satisfaction and reduce turnover.

Dynamic data visualization further enhances the application's functionality, providing HR professionals with real-time insights through interactive graphs and charts. This visual analytics toolkit enables swift and informed decision-making by depicting workforce data, performance trends, and recruitment statistics in a clear and concise manner.

In summary, AI-enabled HR Analytics represents a paradigm shift in HR management, offering organizations the tools they need to optimize processes, engage candidates, and retain top talent. By harnessing the power of AI and data analytics, this innovative solution empowers HR professionals to drive organizational success in today's competitive business environment.

1.2 Motivation

The motivation behind embarking on the AI-Enabled HR Analytics project stems from the recognition of these challenges and the pressing need to modernize and streamline our HR procedures. By harnessing the power of artificial intelligence, natural language processing, and predictive analytics, this project aims to revolutionize how we attract, engage, and retain talent within our organization.

One of the key drivers behind this initiative is the desire to enhance the overall candidate experience. In today's highly competitive job market, attracting top talent and providing them with a positive and seamless recruitment experience is crucial. Through the integration of an AI-driven chatbot for candidate interaction and personalized engagement, we aim to elevate the recruitment process to new heights, ensuring that every candidate feels valued and respected throughout their journey with us.

Furthermore, the project seeks to address the inherent biases and limitations associated with traditional HR practices, particularly in the areas of resume screening and candidate evaluation. By leveraging advanced natural language processing techniques, we aim to automate and standardize the resume screening process, thereby reducing bias and ensuring a fair and objective selection criterion for all candidates.

Moreover, the project aims to unlock the power of data-driven decision-making within the HR domain. By harnessing the vast amounts of data available to us, including historical recruitment data, employee performance metrics, and workforce demographics, we can gain valuable insights into our organizational dynamics and identify trends and patterns that may have previously gone unnoticed. This, in turn, empowers HR professionals to make informed decisions, devise effective strategies, and drive positive outcomes for both the organization and its employees.

In summary, the AI-Enabled HR Analytics project represents a strategic investment in the future of our organization. By embracing cutting-edge technologies and innovative approaches to HR management, we aim to create a more efficient, fair, and data-driven HR ecosystem that not only meets the needs of our organization today but also positions us for success in the rapidly evolving landscape of tomorrow.

1.3 Objective

The goals of the AI-Enabled HR Analytics project are multifaceted, aiming to modernize and optimize various aspects of the human resources management lifecycle. Firstly, the project seeks to enhance candidate selection processes through the implementation of an effective resume screening procedure, leveraging AI algorithms to ensure impartiality and efficiency. Additionally, the inclusion of an AI chatbot facilitates seamless communication with candidates, automating interview scheduling and enhancing the overall applicant experience.

Moreover, the project aims to address retention issues proactively by harnessing the power of predictive analytics. By analysing historical data, potential retention challenges can be identified and pre-emptive measures can be taken to improve employee satisfaction and reduce turnover rates. Furthermore, the provision of dynamic data visualization tools empowers HR professionals with insightful knowledge about the workforce, enabling them to make informed decisions and drive strategic initiatives effectively.

Aligned with these overarching goals, the objectives of the project are focused on simplifying hiring procedures to conserve time and resources, enhancing the applicant experience through interactive and automated communication channels, and leveraging sentiment analysis and predictive analytics to improve HR decision-making processes. Additionally, the project aims to promote proactive retention measures founded on data-driven insights and provide HR workers with accessible graphics to facilitate rapid understanding of complex workforce dynamics. Through the pursuit of these objectives, the AI-Enabled HR Analytics project aims to revolutionize HR management practices and drive organizational success in today's competitive business landscape.

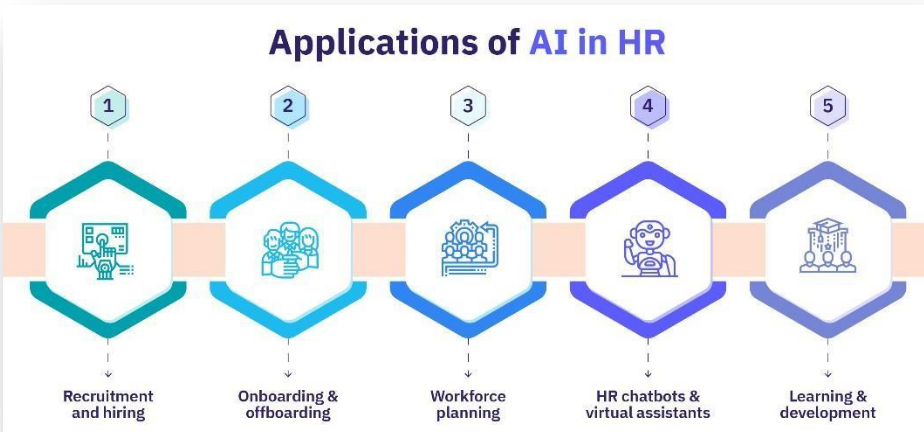


Fig 1. Applications of AI in HR

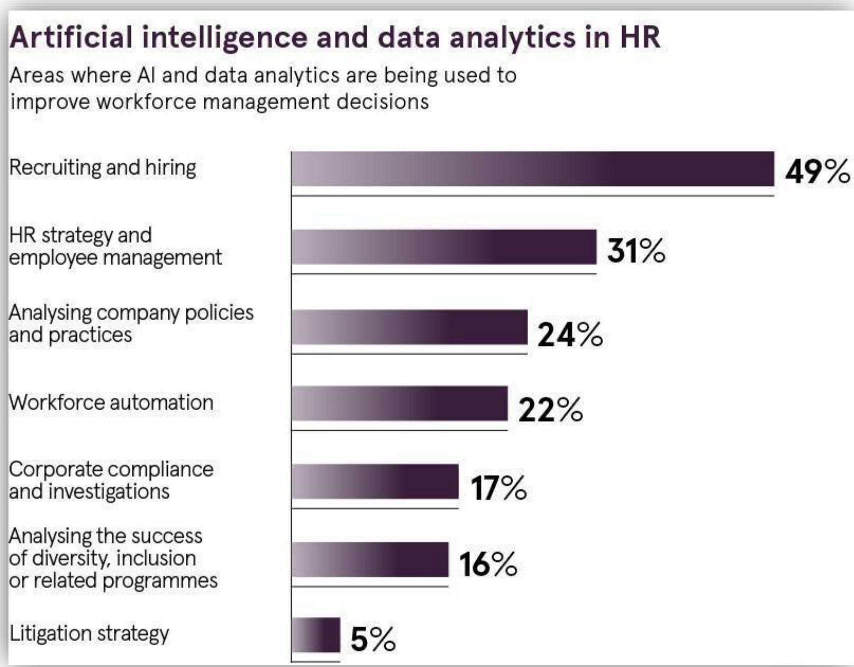


Fig 2. Artificial Intelligence in HR Analytics

Thus, I developed and implemented an AI-enabled HR analytics system aimed at optimizing talent management processes, enhancing employee satisfaction, and improving organizational efficiency through data-driven insights and predictive analytics.

1.4 Problem Statement

In today's rapidly evolving business landscape, human resource (HR) departments are confronted with multifaceted challenges that hinder their ability to effectively manage talent acquisition, engagement, and retention. One significant obstacle lies in the outdated and inefficient processes of resume screening. Traditional methods, reliant on manual review, are not only time-consuming but also susceptible to unconscious biases that may inadvertently exclude qualified candidates from consideration. Moreover, the sheer volume of resumes received further exacerbates this challenge, often resulting in a bottleneck in the recruitment pipeline. Addressing this issue requires a transformative approach that leverages modern technologies, such as artificial intelligence (AI) and natural language processing (NLP), to automate and streamline the screening process while ensuring fairness and impartiality.

Another pressing concern for HR professionals is the need to enhance candidate engagement throughout the recruitment process. Conventional methods of communication and interview scheduling are often disjointed and impersonal, leading to frustration among both candidates and HR personnel. By integrating AI-driven chatbots into the recruitment workflow, organizations can provide candidates with seamless and personalized interactions, from initial application to interview scheduling. This not only improves the candidate experience but also frees up valuable time for HR professionals to focus on more strategic tasks.

The HR department can benefit significantly from the implementation of a resume and cover letter screening application, as manually reviewing resumes and cover letters can be a laborious task. With this tool, HR professionals can efficiently assess the cultural fit of candidates by extracting relevant information from cover letters. Additionally, the application enables streamlined resume screening, allowing HR to extract essential details such as the candidate's name, contact information (including mobile number and email address), field of expertise, years of experience, and skill set.

Employee retention is another critical area where traditional HR practices often fall short. Reactive approaches that only address retention issues after they have escalated can lead to increased turnover rates and decreased morale within organizations. To address this challenge, HR departments need access to tools that can analyze historical data and identify potential retention issues before they arise. By leveraging predictive analytics, organizations can proactively implement strategies to enhance employee satisfaction and loyalty, thereby reducing turnover and preserving organizational stability.

Lastly, effective decision-making within HR departments relies heavily on the ability to extract actionable insights from vast amounts of data. However, traditional reporting methods often present challenges in synthesizing complex information into digestible formats. By implementing dynamic data visualization tools, HR professionals can transform raw data into visually appealing and informative charts and graphs in real time. This enables them to identify trends, patterns, and outliers more efficiently, thereby empowering them to make data-driven decisions that drive organizational success.

1.5 Approach

In the development of my comprehensive HR Analytics project, I meticulously crafted an end-to-end solution catering to the diverse needs of five key end users: Admin, Candidate, HR, Employee, and AVP. Through a combination of CSS styling, HTML for frontend structuring, MongoDB for database and Flask and Python for backend functionality, I created a robust platform that offers seamless navigation and efficient operations.

The incorporation of CSS styling alongside HTML ensured a visually appealing and user-friendly interface, elevating the overall user experience. By leveraging Flask and Python for backend processing, I established a reliable foundation that facilitated smooth data management and processing, ensuring responsiveness and scalability.

My project's user-centric approach acknowledges the unique roles and responsibilities of each end user within the HR ecosystem. From Admins overseeing system management to Candidates navigating the application process, HR professionals conducting resume screenings, Employees accessing HR-related information, and AVPs seeking insightful analytics, my platform caters to the specific needs of every stakeholder.

Through meticulous attention to detail and a thorough understanding of user requirements, I designed a solution that prioritizes functionality, usability, and efficiency. By considering the perspectives and expectations of all end users, I ensured that my platform delivers value across the board.

Each module includes a standard login page and a page for changing passwords. The login page provides an option for users to specify whether they are a candidate or an employee, and the subsequent page is rendered accordingly.

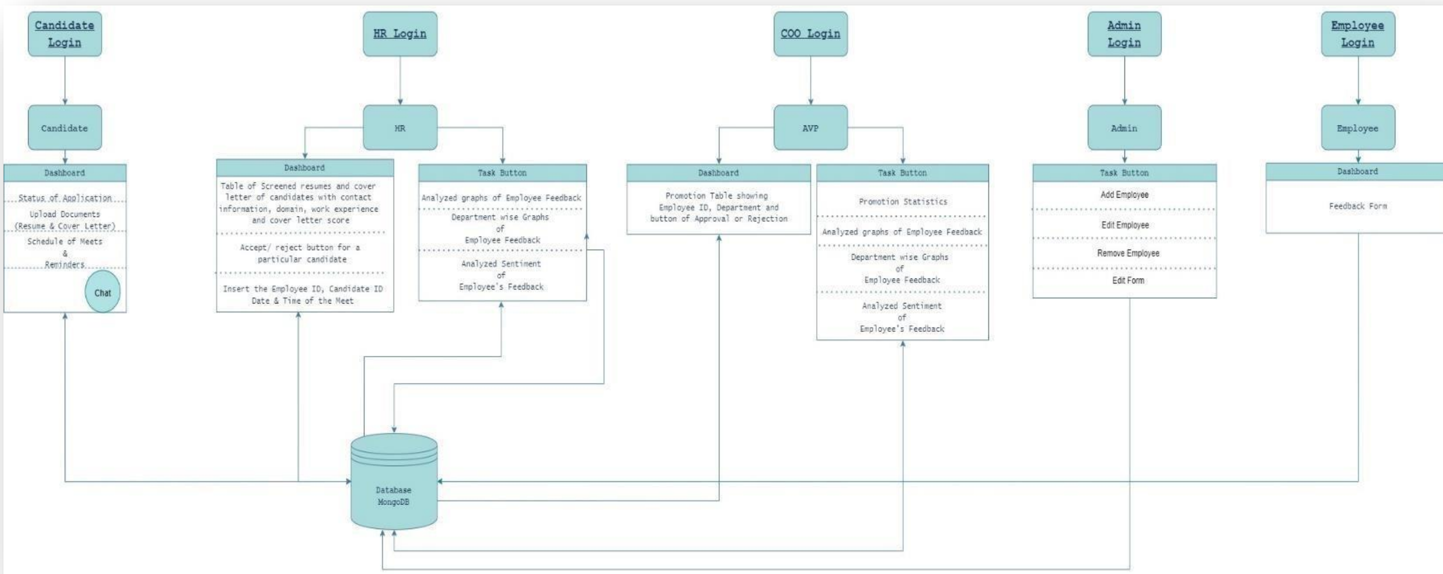


Fig 3. Approach of the project

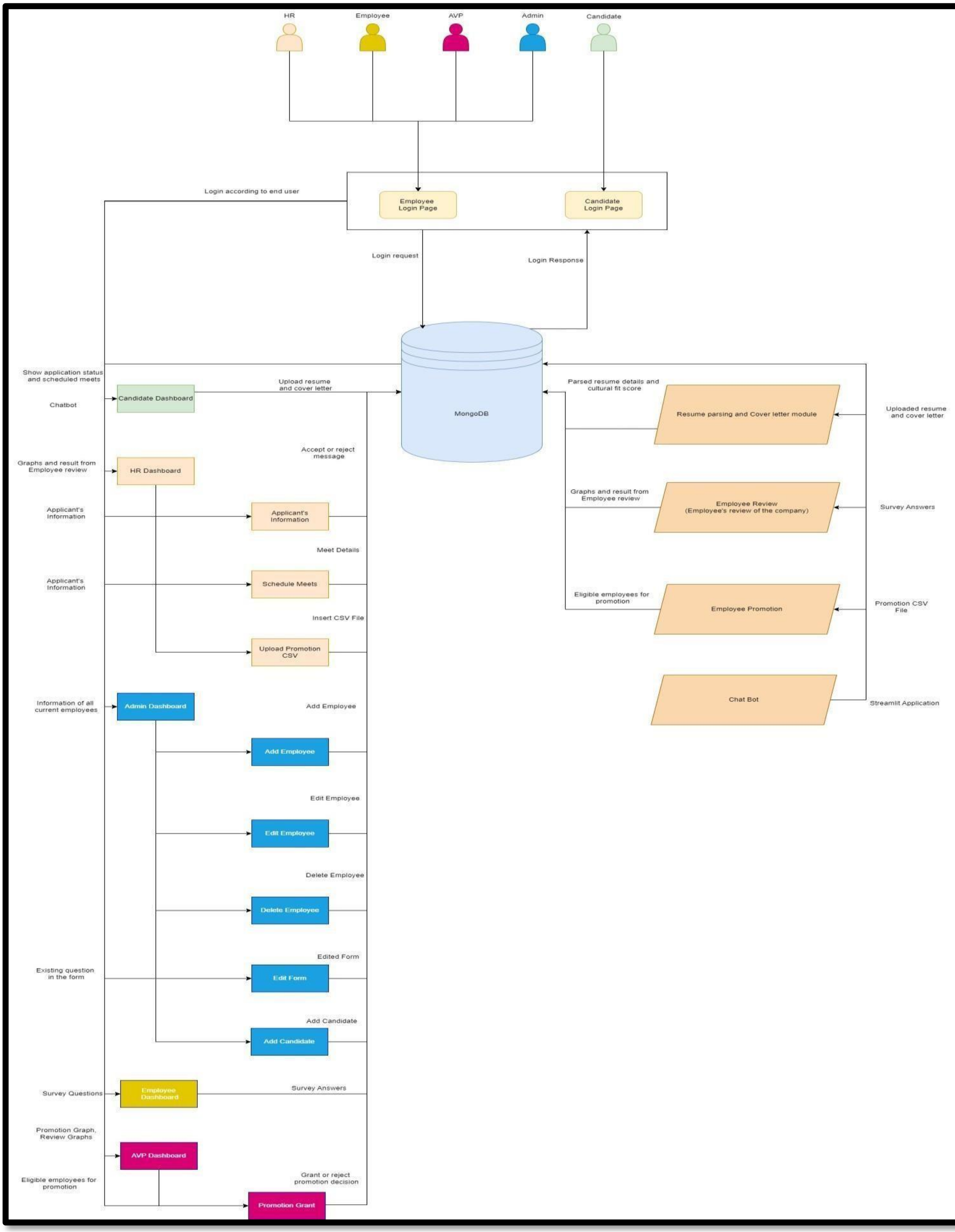


Fig 4. Overview of the project

1.6 Scope of the Project

In the realm of HR analytics, my focus is on two primary domains: employee turnover and recruitment. Leveraging machine learning methodologies such as sentiment analysis and natural language processing (NLP), my approach encompasses a diverse array of applications aimed at enhancing HR processes and decision-making.

Recruitment Process:

- *Resume & Cover Letter Screening*: Employing sentiment analysis and NLP techniques, I meticulously review resumes to extract pertinent details and assess candidate suitability for various positions. By analyzing sentiment, I ensure a more comprehensive evaluation of applicants while streamlining the hiring process. Additionally, sentiment analysis is applied to scrutinize cover letters, enabling me to gauge cultural fit and identify potential red flags. Prescriptive analysis may complement these efforts, optimizing recruitment channels and refining selection criteria based on historical data and performance indicators.
- *Chatbot for Scheduling Interviews and Answering Candidate Questions*: Integration of a chatbot enhances applicant experience and reduces administrative burden by facilitating streamlined communication and scheduling processes.

Turnover of Employees:

- *Employee Evaluations and Feedback*: NLP plays a crucial role in analyzing themes and sentiments within employee evaluations and feedback. This analysis provides insights into prevalent issues, areas of satisfaction, and opportunities for organizational development.
- *Employee Turnover and Leadership Development*: Leveraging machine learning techniques, I gain comprehensive insights into meeting participant and employee satisfaction, thereby obtaining a holistic view of engagement levels. Prescriptive models aid in career development initiatives, mitigating turnover rates, and identifying high-potential employees for leadership roles.

By employing these methodologies and approaches, I aim to enhance HR practices, optimize recruitment processes, and foster organizational development, ultimately contributing to the overall success and growth of the company.

Chapter 2

Literature Review

2.1 Paper 1: A Machine Learning approach for automation of Resume Recommendation system

1. Abstract/Introduction: Recruiting suitable candidates for job vacancies is challenging, particularly in India's vast and dynamic job market with high turnover rates. Manual resume screening is time-consuming and prone to errors due to diverse resume formats. To address this, we propose an automated Machine Learning model. It categorizes resumes based on job descriptions and recommends top candidates to HR.
2. Approach: The objective of this study is to develop a machine learning-based solution to identify suitable candidate resumes from a pool of resumes. The proposed model operates in two main steps: preparation and deployment/inference. The dataset, sourced from online portals and Kaggle, consists of columns for ID, Category (industry sector), and Resume. Preprocessing involves cleansing the resumes by removing special characters, numbers, single-letter words, and stop words. Additionally, stemming and lemmatization are performed to reduce words to their root forms. Feature extraction is then conducted using Term Frequency-Inverse Document Frequency (Tf-Idf). Four classification models—Random Forest (RF), Multinomial Naive Bayes (NB), Logistic Regression (LR), and Linear Support Vector Classifier (Linear SVM)—are applied to predict resume categories. The content-based recommender matches job descriptions with resumes using cosine similarity, while k-Nearest Neighbors identifies resumes closest to the provided job description. The genism library is utilized to standardize the scale of job descriptions and resumes. This approach enhances the efficiency and accuracy of candidate resume selection by leveraging machine learning techniques and document similarity identification methods.
3. Conclusion: An automated machine learning model recommends suitable candidate resumes based on job descriptions, addressing the manual and time-consuming process of resume classification. Achieving 78.53% accuracy with Linear SVM, it suggests further improvement through deep learning models and industry-specific adaptations, with input from HR professionals for iterative refinement.

2.2 Paper 2: Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models using a Framework with OpenAI, Lang chain, and Streamlit

1. Abstract/Introduction: This study introduces a robust framework for building personalized chatbots powered by advanced language models (LLMs) like OpenAI's GPT. By leveraging technologies such as LangChain and Streamlit, developers can create chatbots adept at summarizing documents and answering user queries. The framework addresses the challenge of information overload by efficiently extracting insights from large volumes of text. LLMs, particularly OpenAI's GPT models, play a central role in enabling chatbots to understand and generate natural language. The integration of LangChain enhances linguistic processing capabilities, while Streamlit facilitates user-friendly interfaces. This comprehensive architecture supports the seamless development and deployment of chatbots tailored for document summarization and question-answering tasks.
2. Approach: OpenAI pioneers AI advancements, notably in NLP, robotics, and reinforcement learning. Their focus is on developing AI systems with human-like intelligence, exemplified by the GPT series for natural language tasks. Their proposed web app uses Streamlit, LangChain, and OpenAI APIs to summarize PDFs. Initially, PDFs are chunked and converted into embeddings stored in a vector database. LangChain's OpenAI embeddings class aids in this. When users query, semantic search retrieves relevant chunks. OpenAI API then generates responses based on the query. LangChain and Streamlit complement OpenAI, simplifying NLP tasks and app development, respectively, fostering efficient workflows and professional applications.
3. Conclusion: This study provides a detailed exploration of personalized chatbot development, leveraging large language models (LLMs) for question answering and document summarization. Through the integration of Streamlit, LangChain, and OpenAI technologies, the framework effectively manages information overload by enabling insights extraction from documents. The combination of advanced language models, efficient NLP processing, and user-friendly interface design offers a flexible solution for utilizing LLMs in text-based tasks.

2.3 Paper 3: Sentiment classification for employees reviews using regression vector-stochastic gradient descent classifier (RV-SGDC)

1. Abstract/Introduction: The satisfaction of employees is very important for any organization to make sufficient progress in production and to achieve its goals. Organizations try to keep their employees satisfied by making their policies according to employees' demands which help to create a good environment for the collective. This research delves into sentiment analysis of employee reviews from major companies using a hybrid approach combining lexicon-based and machine learning methods. It introduces a novel Regression Vector-Stochastic Gradient Descent Classifier (RV-SGDC) for sentiment classification. By utilizing sentiment lexicons and supervised learning, the study aims to accurately categorize reviews as positive or negative.
2. Approach: Preprocessing steps are applied to clean the data, including tokenization, lowercase conversion, spelling correction, and removal of numeric values, unnecessary data, punctuation, and stop-words. Sentiment labeling is done using TextBlob to categorize reviews as positive or negative based on polarity scores. Feature extraction involves three techniques: TF-IDF, Bag of Words, and GloVe embeddings. Eight supervised machine learning models are utilized, including Logistic Regression, Random Forest, AdaBoost Classifier, Multilayer Perceptron, Extra Tree Classifier, Support Vector Classifier, Stochastic Gradient Descent, and a novel Regression Vector-Stochastic Gradient Descent Classifier (RV-SGDC), which combines LR, SVC, and SGDC under a majority voting scheme. This comprehensive approach aims to accurately classify employee sentiment from diverse textual data sources, offering insights into job satisfaction and dissatisfaction.
3. Conclusion: This study explores sentiment classification in employee reviews using lexicon and machine learning techniques. Preprocessing prepares the data, and TextBlob labels sentiments. TF-IDF, BoW, and GloVe features train ML models, with RV-SGDC achieving 0.97 accuracy with TF-IDF. Results surpass previous research, highlighting TextBlob's impact and the efficacy of TF-IDF. Statistical tests confirm RV-SGDC's superiority. Future work entails larger datasets and integrating ensemble deep learning models like CNN-LSTM with advanced embeddings for improved accuracy.

Chapter 3

Software Design

I. Admin:

The initial step involves the user inputting login credentials, typically comprising the Employee ID and password, on the login page. Upon successful validation of the admin's credentials, access to the admin dashboard is granted. Within this dashboard, a comprehensive list of current employees within the organization is displayed. The role attributed to the admin in this project primarily revolves around data entry.

The functionalities attributed to the admin encompass various tasks, including the addition of employees and candidates, editing existing employee records, deletion of employee records, and modification of the feedback form utilized for sentiment analysis. When adding an employee, the admin is required to input essential details such as name, email ID, department, position, and contact number. Conversely, the candidate adding system operates on an auto-generated basis. To edit an employee record, the admin must provide the corresponding employee ID requiring modification. In instances where an employee rejoins the company subsequent to departing, the system employs email ID verification to retrieve the employee's previous Emp ID from the database of past collections. This process ensures continuity by assigning the employee their former Emp ID rather than generating a new one.

Notably, the admin holds authority over the modification of the feedback form utilized by employees for providing reviews concerning candidates. This form plays a pivotal role in the sentiment analysis process, facilitating the assessment of employee sentiments towards candidates.

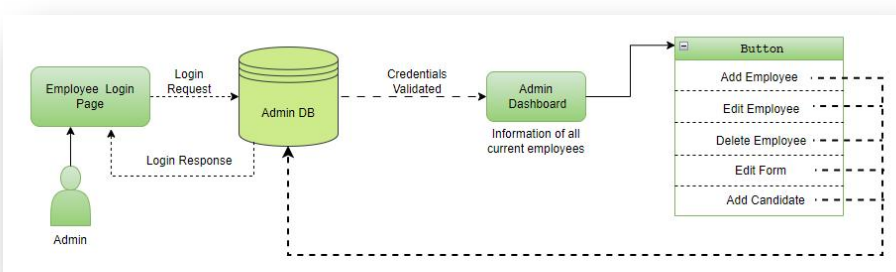


Fig 5. Flow of the Admin

Name of the Collection	Role of that Collection
Login_Details	It stores all the details of the employees like emp_id, name, department, designation, email ID, phone and password
candidate	It stores the details of the candidate namely cand_id, password, contact number, email ID, name and the upload date (of the resume and the cover letter)
past_employees	It stores the details of the employees who have left the company which include emp_id, name, department, designation, email ID and phone
Feedback	It stores all the questions which the admin has edited and they will be displayed on the employee dashboard so as to fill the form

Table 1. Collections used in Admin DB

II. Candidate:

The login process for candidates entails selecting their candidate status on the login page. Upon designation as a candidate, individuals are prompted to input their candidate ID, which commences with 'C', alongside the password conveyed via email by the HR department. Upon successful validation of these credentials, access to the candidate dashboard is granted.

Within the candidate dashboard, pertinent information regarding the status of the application is provided, including details such as the link, date, and time of scheduled meetings for subsequent interview rounds, facilitated by the HR department. Additionally, a designated section is allocated for the mandatory submission of essential documents, namely the resume and cover letter. Once submitted, candidates are unable to resubmit these documents, ensuring data integrity.

Furthermore, candidates are furnished with a chatbot interface designed to address queries pertaining to the company, thereby enhancing communication efficiency and facilitating the candidate's engagement with the recruitment process. This integrated chatbot feature aims to streamline communication channels and provide candidates with timely responses to their inquiries.

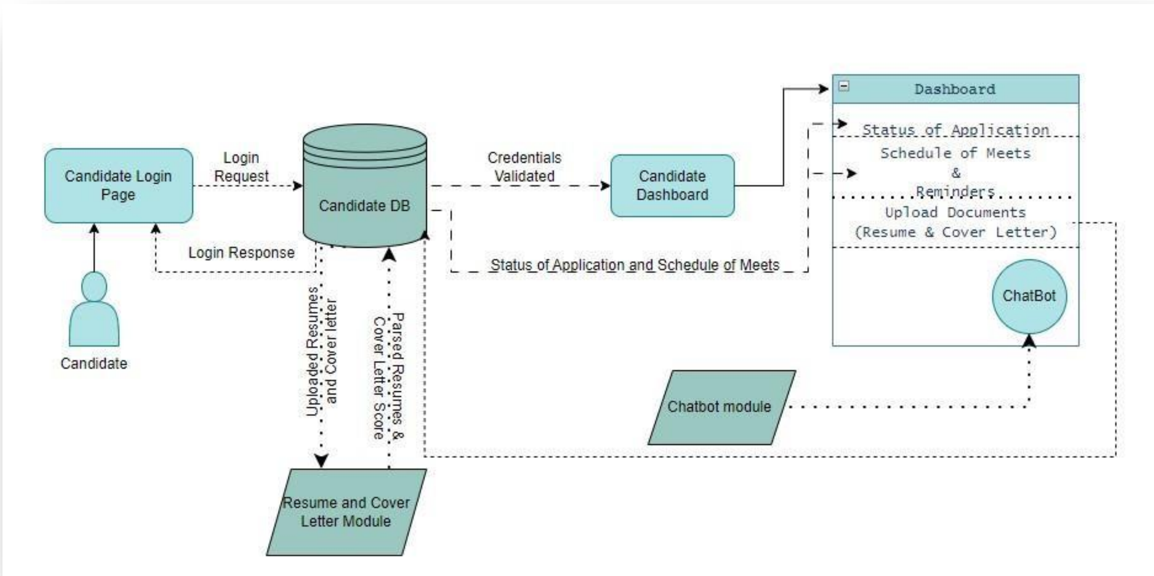


Fig 6. Flow of the Candidate

Name of the Collection	Role of that Collection
candidate	It stores the details of the candidate namely cand_id, password, contact number, email ID, name and the upload date (of the resume and the cover letter)
resume	It stores the extracted details of the candidate like name, contact info, email, domain,experience_category,skills,cultural_fit score, and if the HR has posted a decision then the status of the result
rejected_candidate	If the candidate is rejected then the cand_id along with the reason of rejection will be shown this collection

Table 2. Collections used in Candidate DB

III. Employee:

Upon accessing the login page of the HR Analytics system, authorized employees are prompted to enter their unique Employee ID and password for authentication purposes. Upon successful authentication, the system redirects the authenticated employee to their designated employee page. Implemented within the system is a scheduled task managed by a cronjob, set to execute every six months. This automated task triggers the display of a feedback form to the logged-in employee. The primary objective of this periodic feedback mechanism is to gauge the satisfaction levels of employees within the organization.

The feedback form serves as a pivotal channel for employees to express their sentiments, providing valuable insights into their experiences and perceptions regarding various aspects of their employment within the company. By diligently collecting and analyzing this feedback, HR personnel and higher authorities can gain a comprehensive understanding of employee satisfaction levels and identify potential areas for improvement within the organization.

Name of the Collection	Role of that Collection
Login_Details	It stores all the details of the employees like emp_id, name, department, designation, email ID, phone and password
Feedback	It stores all the questions which the admin has edited and they will be displayed on the employee dashboard so as to fill the form
Feedback_answers	It stores the answers and ratings given by each employee in the feedback form and all of it would be shown as anonymous

Table 3. Collections used in Employee DB

IV. AVP:

The AVP Dashboard serves as a pivotal interface within the HR Analytics system, offering the Assistant Vice President (AVP) a comprehensive overview of various aspects related to employee promotion and feedback analysis.

Upon accessing the AVP Dashboard, the AVP is prompted to input their Employee ID and password for authentication. Upon successful authentication, the AVP gains access to a structured table displaying pertinent information regarding suggested employee promotions. This table includes key details such as Employee ID, department, and corresponding accept and reject buttons for each suggested promotion.

The suggestions for employee promotion are generated by a Python code leveraging a Gradient Boosting Classifier trained on a myriad of parameters. These parameters encompass vital metrics such as previous year ratings, length of service, Key Performance Indicators (KPIs) met, awards won, and average training scores. The AVP is entrusted with the responsibility of evaluating and deciding upon the promotion of employees based on the information provided in a CSV file uploaded by the AVP, containing the aforementioned parameters.

Furthermore, the AVP is presented with insightful statistics in the form of graphical representations, facilitating a deeper understanding of employee promotion dynamics. These graphs encompass various dimensions such as KPIs met, awards won, departmental distribution, length of service categorization, total score amalgamating training metrics, and service tenure categories.

In addition to promotion statistics, the AVP is furnished with comprehensive feedback analysis, derived from a Python code employing the RVSGDC algorithm. This algorithm, a combination of Logistic Regression, Support Vector Machine, and Stochastic Gradient Descent supported by hard voting, is trained on a feedback dataset and solicits inputs from feedback responses.

The feedback statistics are portrayed graphically, offering insights into overall and department-wise ratings across distinct categories, including overall rating, work-life balance, cultural values, career opportunities, company benefits, and senior management perception. Additionally, sentiment analysis is conducted on the feedback responses, providing a breakdown of sentiments into positive, neutral, and negative categories.

Name of the Collection	Role of that Collection
Login_Details	It stores all the details of the employees like emp_id, name, department, designation, email ID, phone and password

Employee_Promotion	It stores the data from the CSV file which is uploaded by the AVP
Promotion_upload date	It stores the upload date of the CSV file in the form of timestamp
Predicted_Promotion	It stores the employee ID and the predicted promotion status that is promoted and not promoted
Plots_Promotion	It stores the plots images in the form of binary format along with the plot ID.
Feedback_answers	It stores the answers and ratings given by each employee in the feedback form and all of it would be shown as anonymous
Plots_Review	It stores the plots images in the form of binary format along with the plot ID.

Table 4. Collections used in AVP DB

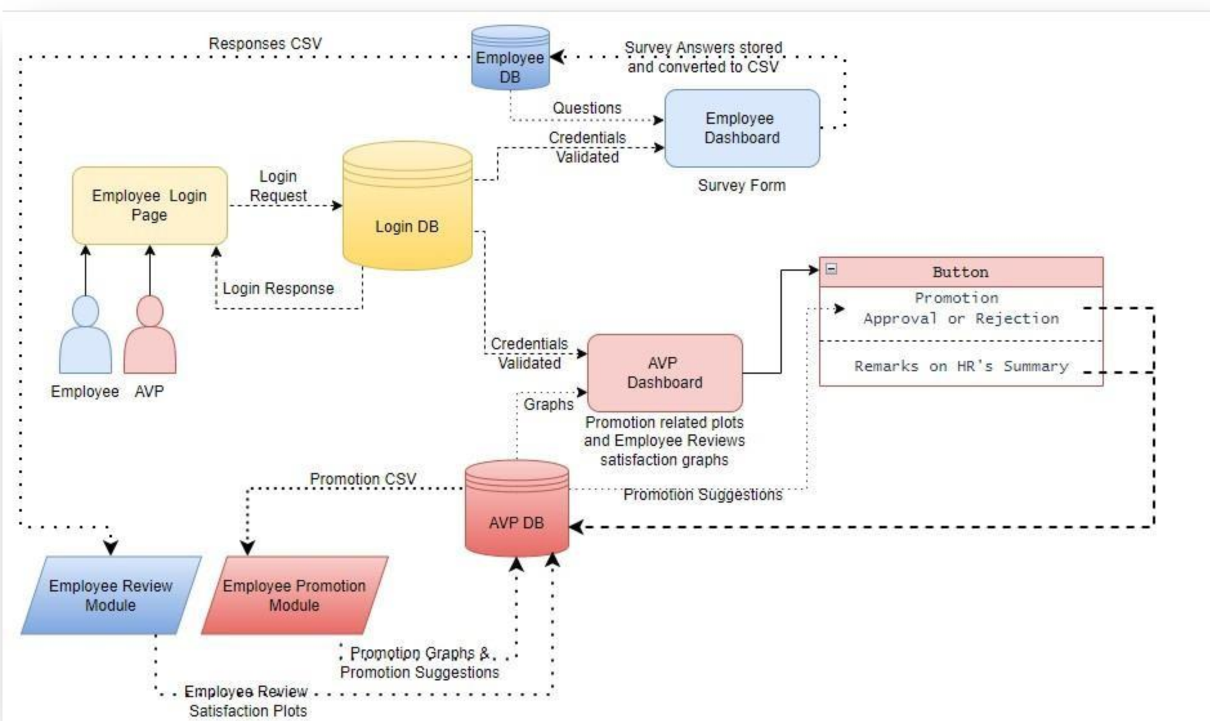


Fig 7. Flow of the Employee & AVP

V. HR:

The HR Dashboard serves as a pivotal interface within the HR Analytics system, offering Human Resources personnel a comprehensive toolkit for streamlined resume screening and feedback analysis.

Upon accessing the HR Dashboard, HR personnel are prompted to input their Employee ID and password for authentication. Upon successful authentication, HR personnel gain access to a structured table presenting pertinent information regarding screened resumes and associated cover letter cultural fit scores.

The table showcases essential details extracted from resumes, including a downloadable link to the resume PDF, candidate name, contact information (comprising mobile number and email address), domain mastery as determined by a Naive Bayes model trained on a sample dataset, categorized work experience denoted in Fresher, Beginner, Mid-level, and Experienced brackets, skills mapped via Natural Language Processing (NLP) techniques utilizing a skills dataset, and a downloadable link to the cover letter PDF.

The cover letter cultural fit score is calculated leveraging sentence transformers and cosine similarity metrics, juxtaposing the cover letter content against the company's mission and vision to ascertain alignment and cultural compatibility.

Moreover, the HR Dashboard features intuitive functionalities aimed at enhancing the recruitment process. For each candidate listed in the table, HR personnel are equipped with buttons to seamlessly accept or reject the candidate, streamlining the decision-making process.

Furthermore, HR personnel are empowered with the ability to schedule meetings with candidates directly from the dashboard. A dedicated "Schedule Meeting" button accompanies each candidate entry, allowing HR personnel to efficiently manage interview logistics. Upon clicking this button, a new tab opens, presenting HR personnel with a form to input essential meeting details. This form prompts HR personnel to upload the meeting link, specify the meeting date and time, and provide a brief description regarding the meeting agenda.

In addition to resume screening functionalities, the HR Dashboard furnishes HR personnel with insightful feedback statistics gleaned from a Python code employing the RVSGDC algorithm. This algorithm, a fusion of Logistic Regression, Support Vector Machine, and Stochastic Gradient Descent supported by hard voting, is trained on a feedback dataset and solicits inputs from feedback responses.

The feedback statistics are visually represented through graphs, offering HR personnel insights into overall and department-wise ratings across distinct categories, encompassing overall rating, work-life balance, cultural values, career opportunities, company benefits, and senior management perceptions. Furthermore, sentiment analysis is conducted on the feedback responses, categorizing sentiments into positive, neutral, and negative categories.

Name of the Collection	Role of that Collection
Login_Details	It stores all the details of the employees like emp_id, name, department, designation, email ID, phone and password
resume	It stores the extracted details of the candidate like name, contact info, email, domain, experience_category, skills, cultural_fit score, scheduled meeting details, and if the HR has posted a decision then the status of the result
rejected_candidate	If the candidate is rejected then the cand_id along with the reason of rejection will be shown this collection
Feedback_answers	It stores the answers and ratings given by each employee in the feedback form and all of it would be shown as anonymous
Plots_Review	It stores the plots images in the form of binary format along with the plot ID.

Table 5. Collections used in HR DB

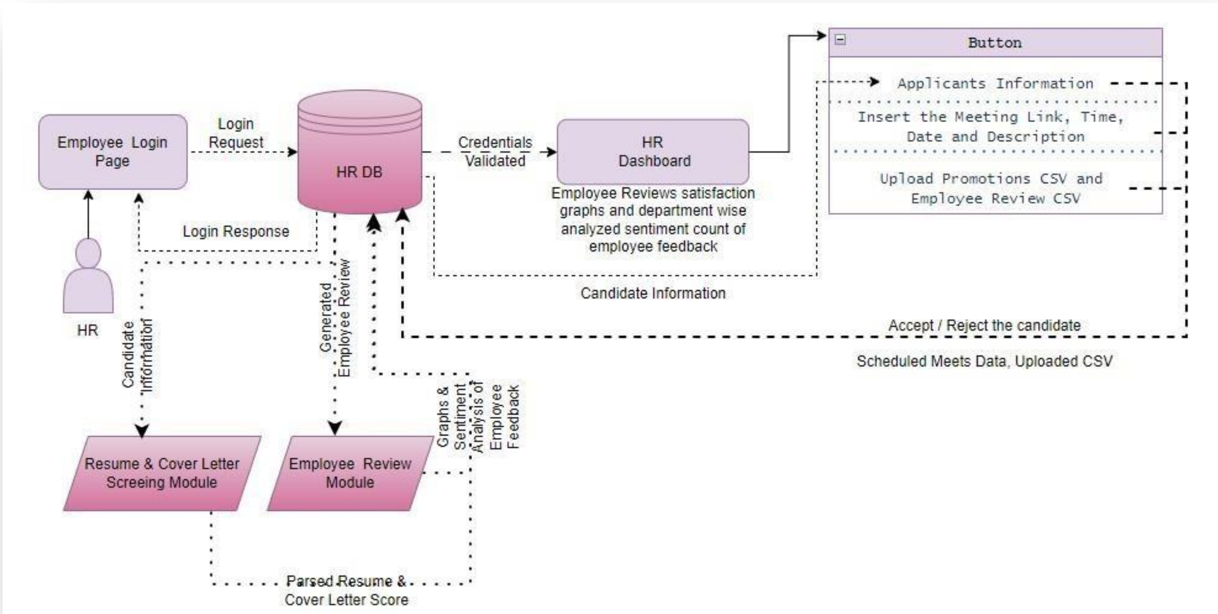


Fig 8. Flow of the HR

Chapter 4

Experimental Simulations and Results

1.1 Model Development

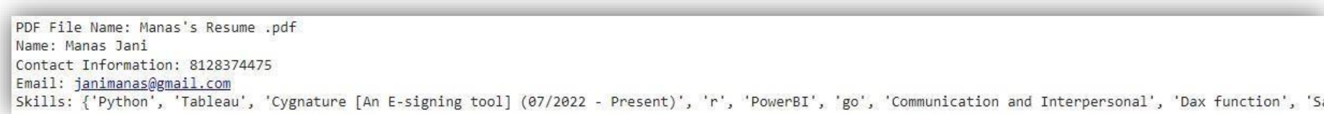
I. Resume & Cover Letter Screening:

1. Reading the skills Dataset

	SKILLS
0	Experiment Design
1	Prototyping
2	Product Testing
3	Intellectual Property Management
4	Patent Analysis

Fig 9. Skills Dataset

2. In preprocessing the skills dataset, text is converted to lowercase for consistency. A function is crafted to extract skills, considering bigrams and trigrams to capture multi-word skills. Another function is designed to extract candidate names, utilizing pattern matching or named entity recognition. Additionally, a function is created to extract contact information like email addresses and phone numbers using regular expressions. These steps ensure the dataset is well-prepared for applications like resume screening, aiding in efficient talent acquisition processes. Lastly, a random PDF is given and the information is extracted by calling the above-mentioned functions.



```
PDF File Name: Manas's Resume .pdf
Name: Manas Jani
Contact Information: 8128374475
Email: janimanas@gmail.com
Skills: {'Python', 'Tableau', 'Cygnature [An E-signing tool] (07/2022 - Present)', 'r', 'PowerBI', 'go', 'Communication and Interpersonal', 'Dax function', 'Sa
```

Fig 10. Extracted Data

3. For Domain extraction we preprocess the extracted resume text by applying lowercasing, removing blank spaces, removing links, removing alpha-numeric values, removing stop words, performing lemmatization (to extract only the stem of the word) and finally tokenize the extractions.

```python
    text = text.lower()
    text = re.sub(r"\s+", " ", text).strip()
    text = re.sub(r"https?://\S+|www\.\S+", " ", text)
    text = re.sub(r"<.*?>|&([a-z0-9]+|#[0-9]{1,6}|#x[0-9a-f]{1,6});", " ", text)
    text = re.sub(r"\b(?:\d{3}[-.\s]??\d{3}[-.\s]??\d{4}|\(\d{3}\)\s*\d{3}[-.\s]??\d{4}|\d{3}[-.\s]??\d{4})\b", " ", text)
    text = text.translate(str.maketrans('', '', string.punctuation))
    text = re.sub(r"\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-Z|a-z]{2,}\b", " ", text)
    text = re.sub(r"[^a-zA-Z\s]", " ", text)
    tokens = word_tokenize(text)
    tokens = [word for word in tokens if word not in stop_words]
    tokens = [lemmatizer.lemmatize(word) for word in tokens]
    preprocessed_text = ' '.join(tokens)
```

Fig 11. Preprocessing Resume Text

4. Training different models by defining the pipelines of TF-IDF vectorization, standard scaling (if required) and the model name. We compare 5 models namely Random Forest, Support Vector Machine, Multinomial Naïve Bayes, Logistic Regression and Neural Network. I had use autoML to determine the best parameter of each model.

```
est parameters: {'clf__max_depth': 20, 'clf__min_samples_leaf': 1, 'clf__min_s
est cross-validation accuracy: 0.9903902964584854
est loss: 0.5800176092115225
est accuracy: 1.0
lassification Report:
                      precision    recall  f1-score   support

             Advocate       1.00      1.00      1.00         4
                 Arts       1.00      1.00      1.00         4
   Automation Testing       1.00      1.00      1.00         4
           Blockchain       1.00      1.00      1.00         4
     Business Analyst       1.00      1.00      1.00         4
       Civil Engineer       1.00      1.00      1.00         4
         Data Analyst       1.00      1.00      1.00         4
         Data Science       1.00      1.00      1.00         4
             Database       1.00      1.00      1.00         4
      DevOps Engineer       1.00      1.00      1.00         4
     DotNet Developer       1.00      1.00      1.00         4
        ETL Developer       1.00      1.00      1.00         4
Electrical Engineering       1.00      1.00      1.00         4
                   HR       1.00      1.00      1.00         4
```

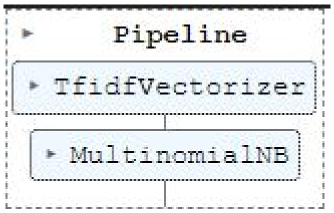
```
est cross-validation accuracy: 0.983178674451743
est loss: 0.2357708639841659
est accuracy: 0.9903846153846154
lassification Report:
                      precision    recall  f1-score   support

             Advocate       1.00      1.00      1.00         4
                 Arts       1.00      1.00      1.00         4
   Automation Testing       1.00      1.00      1.00         4
           Blockchain       1.00      1.00      1.00         4
     Business Analyst       1.00      1.00      1.00         4
       Civil Engineer       1.00      1.00      1.00         4
         Data Analyst       1.00      1.00      1.00         4
         Data Science       0.80      1.00      0.89         4
             Database       1.00      0.75      0.86         4
      DevOps Engineer       1.00      1.00      1.00         4
     DotNet Developer       1.00      1.00      1.00         4
        ETL Developer       1.00      1.00      1.00         4
Electrical Engineering       1.00      1.00      1.00         4
                   HR       1.00      1.00      1.00         4
               Hadoop       1.00      1.00      1.00         4
```

```
est parameters: {'clf__C': 10, 'clf__gamma': 'scale', 'clf__kernel': 'linear'}
est cross-validation accuracy: 0.9855767559865152
est accuracy: 0.9903846153846154
lassification Report:
                      precision    recall  f1-score   support

             Advocate       1.00      1.00      1.00         4
                 Arts       1.00      1.00      1.00         4
   Automation Testing       1.00      1.00      1.00         4
           Blockchain       1.00      1.00      1.00         4
     Business Analyst       1.00      1.00      1.00         4
       Civil Engineer       1.00      1.00      1.00         4
         Data Analyst       1.00      1.00      1.00         4
         Data Science       0.80      1.00      0.89         4
             Database       1.00      0.75      0.86         4
      DevOps Engineer       1.00      1.00      1.00         4
     DotNet Developer       1.00      1.00      1.00         4
        ETL Developer       1.00      1.00      1.00         4
Electrical Engineering       1.00      1.00      1.00         4
                   HR       1.00      1.00      1.00         4
               Hadoop       1.00      1.00      1.00         4
```

```
est parameters: {'clf__alpha': 0.1}
est cross-validation accuracy: 0.9687554304382581
est loss: 0.07186026848159627
est accuracy: 0.9903846153846154
lassification Report:
                      precision    recall  f1-score   support

             Advocate       1.00      1.00      1.00         4
                 Arts       1.00      1.00      1.00         4
   Automation Testing       1.00      1.00      1.00         4
           Blockchain       1.00      1.00      1.00         4
     Business Analyst       1.00      1.00      1.00         4
       Civil Engineer       1.00      1.00      1.00         4
         Data Analyst       1.00      1.00      1.00         4
         Data Science       1.00      1.00      1.00         4
             Database       1.00      1.00      1.00         4
      DevOps Engineer       1.00      1.00      1.00         4
     DotNet Developer       1.00      1.00      1.00         4
        ETL Developer       1.00      1.00      1.00         4
Electrical Engineering       1.00      1.00      1.00         4
                   HR       1.00      1.00      1.00         4
```

Fig 12. Classification Reports of Different Models (Test Data)

5. As the Naïve Bayes model performs the best so we save the model using joblib and thus reading the resume through fitz library and giving the prediction of the domain.




```
pdf_path = r'D:\HR-Analytics-Final\src\uploads\Vaishali_Panchal_CV.pdf'

# Extract text from the PDF
pdf_text = extract_text_from_pdf(pdf_path)

# Preprocess the extracted text
preprocessed_pdf_text = preprocess_text(pdf_text)

# Transform using the fitted TF-IDF vectorizer
#pdf_embeddings = tfidf_vectorizer.transform([preprocessed_pdf_text])

# Make prediction using the loaded model
predicted_domain_number = loaded_model.predict([preprocessed_pdf_text])

# Convert the predicted domain number to domain name
predicted_domain_name = label_encoder.inverse_transform(predicted_domain_number)

print("Predicted Domain:", predicted_domain_name)

Predicted Domain: ['Web Designing']
```

Fig 13. Defining the Pipeline and Predicting the Domain by Calling the Functions

6. To extract work experience we need to preprocess the extracted text by removing the phone numbers

```
def remove_phone_numbers(text):
    # Regular expression pattern to match phone numbers
    phone_pattern = r'\b(?:\d{3}[-.\s]|\(\d{3}\)\s*)\d{3}[-.\s]?\d{4}\b'
    return re.sub(phone_pattern, '', text)
```

Fig 14. Preprocessing by Removing the Phone Number

7. I binned the extracted years and thus predict the experience category

```
def bin_years_of_experience(years):
    if years <= 1:
        return "Fresher"
    elif 2 <= years <= 4:
        return "Beginner"
    elif 5 <= years <= 9:
        return "Mid-level"
    else:
        return "Experienced"

# Example usage
pdf_path = r"D:\HR-Analytics-Final\src\uploads\Auto_Testing.pdf"
years_of_experience = extract_work_experience(pdf_path)
experience_category = bin_years_of_experience(years_of_experience)
#print("Years of work experience:", years_of_experience)
print("Experience category:", experience_category)

xperience category: Beginner
```

Fig 15. Binning and Predicting the Experience Category

8. For the cover letter screening, I will be firstly extracting the text from the PDF and then defining the keywords which would be matched with the pdf so as to determine the score.

```
keywords ="""My work adds business value to customers.
            My work adds career value to fellow employees.
            My work adds social value to communities.
            My work shows my integrity.
            My work shows my team work and team spirit.
            My work shows my trustworthy quality.
            My work showcased digital transformation.
            My work shows my simplicity.
            My work shows my entrepreneurial spirit.
            I have good communication skills.
            I have good interpersonal skills.
            My work shows innovation.
            My work shows resilience.
            I am confident about my work.
            I have leadership qualities.
            My work always meets deadlines."""
```

Fig 16. Defining Keywords

9. Furthermore, I will be preprocessing the extracted text by removing the punctuations, handles, lowercasing the text, removing alpha-numeric data, removing stopwords, stemming and tokenizing. Moreover, I get the output after using sentence transformers.

```
print(tokenize_cover_letter)
print(tokenize_keywords)

[' february 2017 dear braintree, letter express interest posting pm intern role.', 'iõm currently pursuing masters computer
[' work adds business value customers.', 'work adds career value fellow employees.', 'work adds social value communities.',
```

Fig 17. Preprocessed Text

10. Lastly, I applied cosine similarity and torch so as to get the dimensions of the orientation of the matrix and to get the cover letter score on the similarity index.

```
cosine_similarity_score

tensor(0.3290)
```

This similarity score determines that the candidate is a good cultural fit if the score is **greater than 0.4** or else he is not a good fit.

Fig 18. Cover Letter Score

II. Chatbot:

1. The "get response" function is defined using conditional statements, categorizing questions and responses. Categories and subcategories are established to organize content on the chatbot page efficiently. This approach streamlines user interactions by providing relevant information based on their inquiries. Through systematic categorization, the chatbot enhances user experience and facilitates seamless communication.

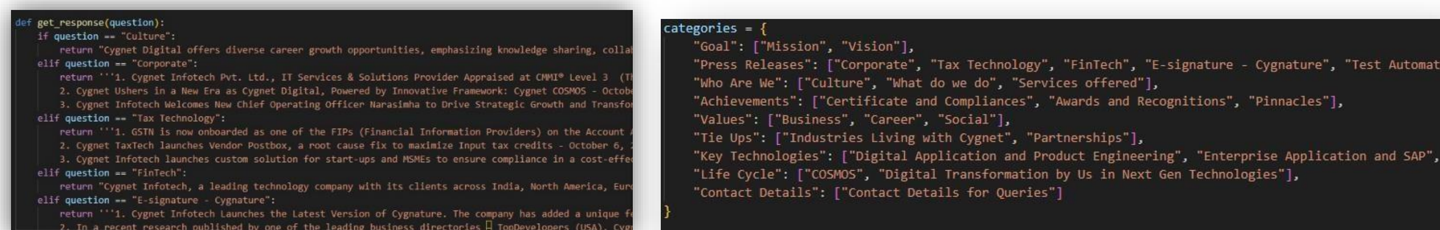


Fig 19. Creating Function of Responses and Defining Categories

2. A Streamlit page is developed to showcase defined categories and subcategories, allowing users to select their preferences. Upon selection, the corresponding response is retrieved. Additionally, HTML and CSS are utilized to customize the page's aesthetics and enhance its visual appeal and usability.

Fig 20. Streamlit Page of the Code

III. Employee Review:

1. Reading the CSV and plotting visual graphs for counting the ratings of the employees in categories namely: Overall Ratings, Work Balance Stars, Culture Value stars, Career Opportunity Stars, Company Benefit Stars and Senior Management Stars.

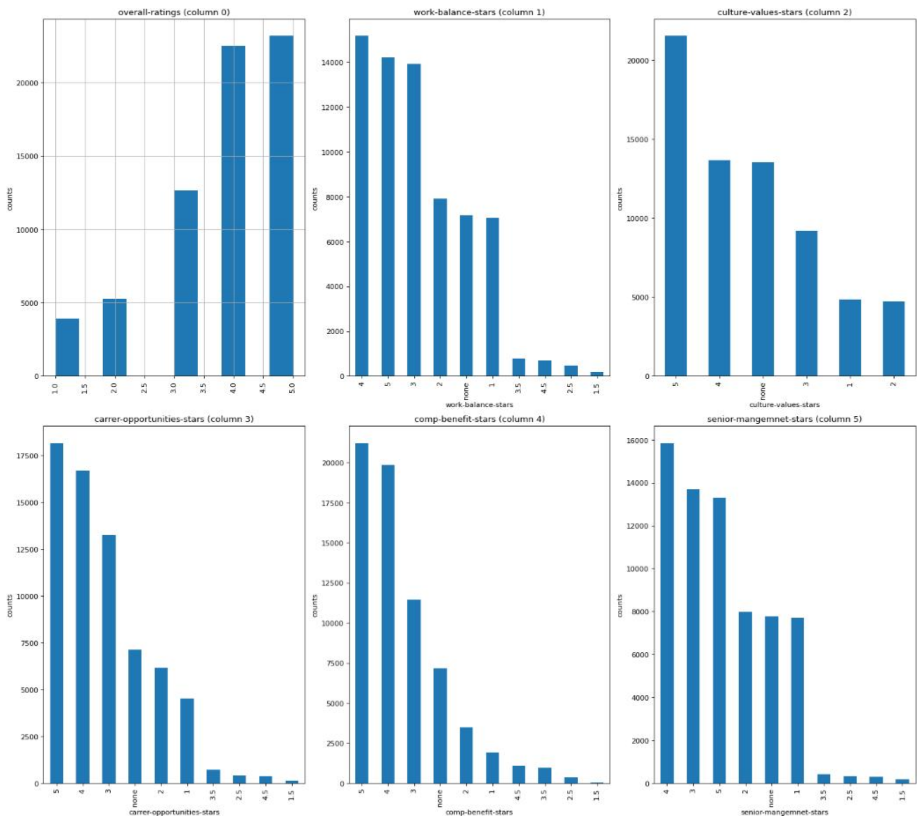


Fig 21. Analysis of the Employee Ratings

2. I merged textual columns (pros, cons, advice to management, and overall review) to create a new column, "Employee Summary," consolidating key insights. Subsequently, the original columns are dropped to streamline data clarity. Furthermore, preprocessing techniques, including removing handles, punctuation, stopwords, alphanumeric characters, lowercasing, tokenization, and lemmatization, are applied to refine the text for analysis, ensuring accurate and efficient natural language processing procedures.

```
0        best compani work forpeopl smart friendlyburea...
1        move speed light burn food food food cafe main...
2        great balanc secur fun project softwar engin a...
3        best place work also ca find compani actual de...
4        uniqu one kind dream jobgoogl world everi comp...
                                ...
67524    enrich experi beginn bad long term wide rang t...
67525    complex interest microsoft chang role either c...
67526    good place worknic place work good atmospher c...
67527    competit work place overload work item grow po...
67528    use greatcompens health benefit brand name rec...
Name: emp_summary, Length: 67529, dtype: object
```

Fig 22. Preprocessed Text of the Merged Indices

3. To label the unlabeled data for sentiment analysis, TextBlob and polarity scores are employed to assess the sentiments of the merged text from "employee_summary." Polarity scores provide insights into the sentiment orientation, aiding in categorizing the data into positive, negative, or neutral sentiments. These labeled sentiments serve as training data for developing sentiment analysis models.

```
1  move speed light burn food food food cafe main...  Positive
2  great balanc secur fun project softwar engin a...  Positive
```

Fig 23. Labelled Data

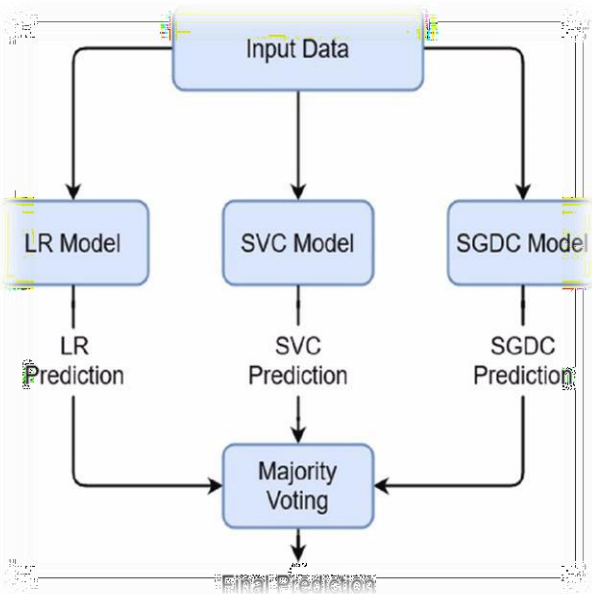
4. Upon data analysis, an imbalance in sentiment distribution was observed, predominantly favoring positive sentiments. To mitigate this imbalance during model training, undersampling was employed. This involved randomly shuffling the data and then reducing the number of instances in the majority class (positive sentiments) to achieve a more balanced distribution. Undersampling ensures that the model is trained effectively across all sentiment categories.

```
sentiment
Positive    52789
Negative     7418
Neutral      7322
Name: count, dtype: int64
```

```
sentiment
Positive    7322
Negative    7322
Neutral     7322
Name: count, dtype: int64
```

Fig 24. Balancing the Imbalanced Data

5. To facilitate text analysis by the machine learning model, the text inputs were transformed into numerical representations using TF-IDF vectorization. Subsequently, a train-test split was established with a ratio of 75:25. A model was defined namely RVSGDC (an amalgamation of Logistic Regression, Support Vector Machine and Stochastic Gradient descent) where the best parameters were determined by hard voting. Furthermore, it was trained on the training data. Evaluation metrics including test accuracy, precision, and recall were computed to assess the model's performance on unseen data.




```
Accuracy: 0.9380917698470502
Precision: 0.9395439795694674
Recall: 0.9380917698470502
```

Fig 25. Model Architecture and Performance on Test Data

6. Lastly, I saved the trained model using joblib and predicted a random string using the model

```
random_string = "I  didnot enjoy the movie. It was bad!"
random_string_tfidf = tfidf_vectorizer.transform([random_string])
predicted_sentiment = voting_clf.predict(random_string_tfidf)
print("Predicted sentiment:", predicted_sentiment[0])

Predicted sentiment: Negative
```

Fig 26. Model (RVSGDC) Prediction

IV. Employee Promotion:

1. Initially, the CSV file was read, and null value analysis was conducted, revealing missing data in the "Previous_Year_Ratings" and "Education" fields. Given that null values in "Previous_Year_Ratings" could be attributed to new employees, they were retained. However, as education information is essential and cannot be null, null entries in the "Education" field were addressed by replacing them with the mode of the domain to which the employee belongs. Subsequently, the preprocessing of the CSV file was completed, ensuring data integrity and usability for further analysis.

	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_se
0	65438	Sales & Marketing	region_7	Master's & above	f	sourcing	1	35	5.0	
1	65141	Operations	region_22	Bachelor's	m	other	1	30	5.0	
2	7513	Sales & Marketing	region_19	Bachelor's	m	sourcing	1	34	3.0	
3	2542	Sales & Marketing	region_23	Bachelor's	m	other	2	39	1.0	
4	48945	Technology	region_26	Bachelor's	m	other	1	45	3.0	

Fig 27. Glimpse of the Data

```
previous_year_rating
3.0    18618
5.0    11741
4.0     9877
1.0     6223
2.0     4225
NaN     4124
me: count, dtype: int64
```

```
Null rating counts of employees with length of service 1
 4124
Null rating counts of employees with length of service 1 and promoted
 339
```

```
Sales & Marketing  :  Bachelor's
Operations  :  Bachelor's
Technology  :  Bachelor's
Analytics  :  Bachelor's
R&D  :  Bachelor's
Procurement  :  Bachelor's
Finance  :  Bachelor's
HR  :  Bachelor's
Legal  :  Bachelor's
```

```
df.education = df.education.fillna(df.education.mode()[0])
```

Fig 28. Analyzing and Filling Null Values

2. Data analysis was conducted to identify factors influencing employee promotion. Features contributing significantly to promotion were determined for model training, focusing solely on these influential factors.

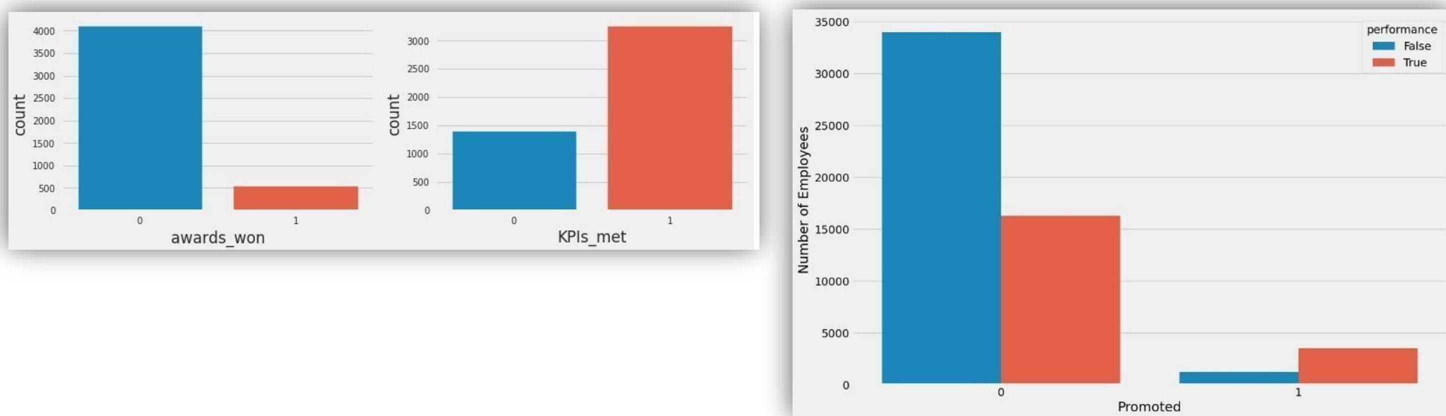


Fig 29. KPIs met and Awards Won, Combined View in Performance Graph

The analysis of Figure 28 indicates that both awards won and KPIs met are significant factors in employee promotion. These metrics are combined in the performance evaluation of promoted and non-promoted employees.

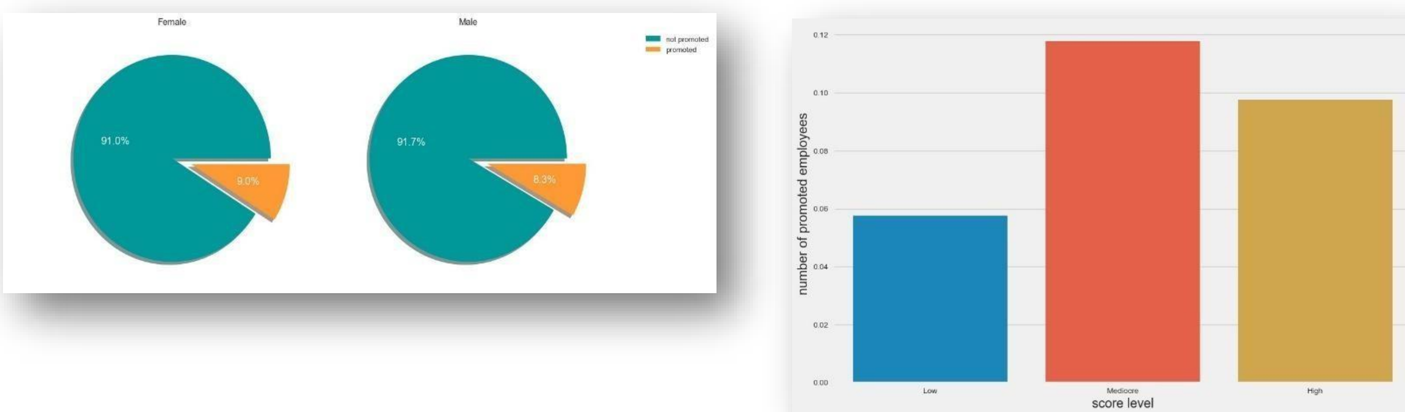
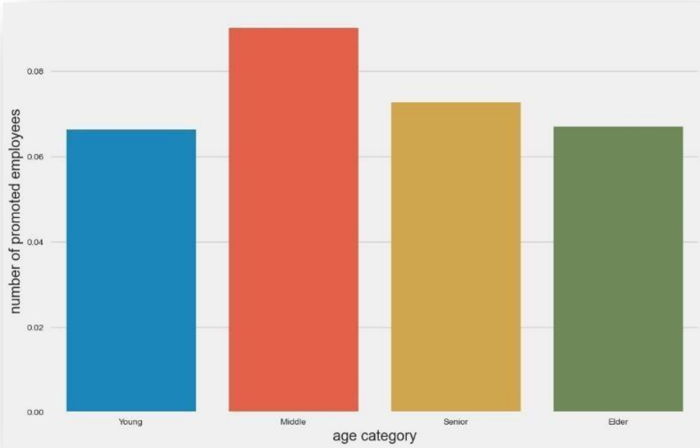


Fig 30. Gender & Total Training Score Graph

Figure 29 analysis reveals gender neutrality in promotion rates, indicating equal promotion likelihood for both genders. Additionally, the total training score, derived from the average training score * number of trainings, emerges as a crucial factor for promotion, with the highest promotion rates observed in the mediocre score category.



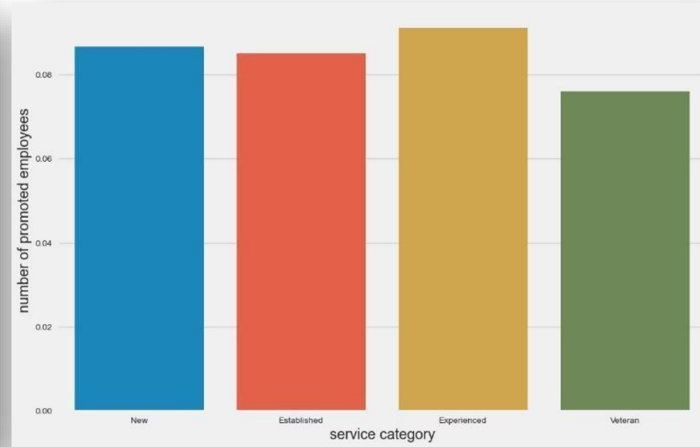
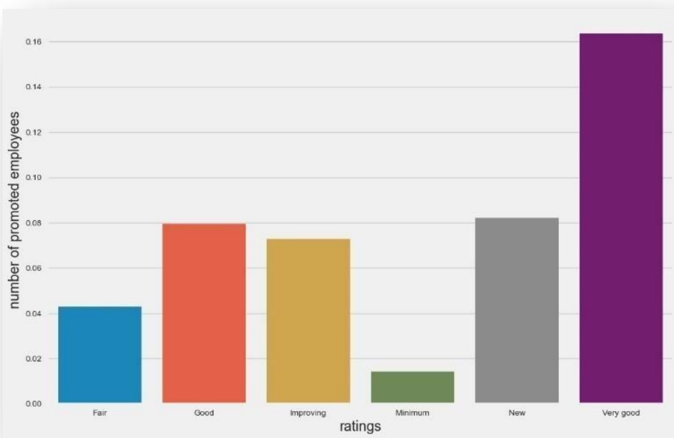


Fig 31. Age Category & Length of Service

Figure 30 analysis suggests that age is a significant but biased factor for promotion and should be omitted from consideration. Conversely, length of service is deemed insignificant for promotion.



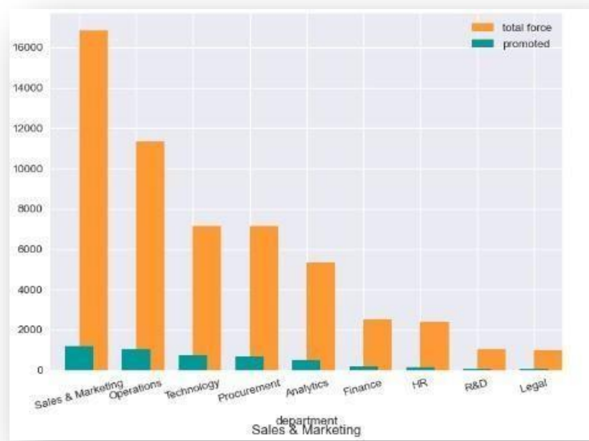


Fig 32. Previous Years Ratings & Department Wise Promotion

Figure 31 analysis underscores the significance of previous year ratings in employee promotion, with a notable preference for those with a "very good" rating. Additionally, the second graph provides supplementary insight, indicating that the Sales & Marketing department has the highest promotion rate, yet it's contingent on department size rather than inherent departmental characteristics. Thus, departmental affiliation alone does not influence promotion likelihood.

3. Before model evaluation, unnecessary columns were removed from the dataset. The dropped columns include department, region, education, gender, recruitment_channel, service category, and age label.

4. To mitigate data imbalance, I applied Min Max scaling and random oversampling. This approach ensures parity between promoted and non-promoted employees. Oversampling was chosen to preserve data integrity and prevent information loss.

```
is_promoted
0    50140
1     4668
Name: count, dtype: int64
```

```
(100280, 7)
is_promoted
0    50140
1    50140
Name: count, dtype: int64
```

Fig 33. Data Balancing

5. The model was trained using a gradient boosting classifier, with an 80:20 train-test split. Subsequently, both train and test accuracies were computed, along with generating a classification report. This methodological approach ensures robust evaluation and validation of the model's predictive performance.

```
0.9738066746443292
0.9566812923813323
```

```
              precision    recall  f1-score   support

           0       0.91      1.00      0.95     11403
           1       1.00      0.92      0.96     13667

    accuracy                           0.96     25070
   macro avg       0.96      0.96      0.96     25070
weighted avg       0.96      0.96      0.96     25070
```

Fig 34. Classification Report of Gradient Boosting Classifier Model

4.2 Web Development

I. Login Page:

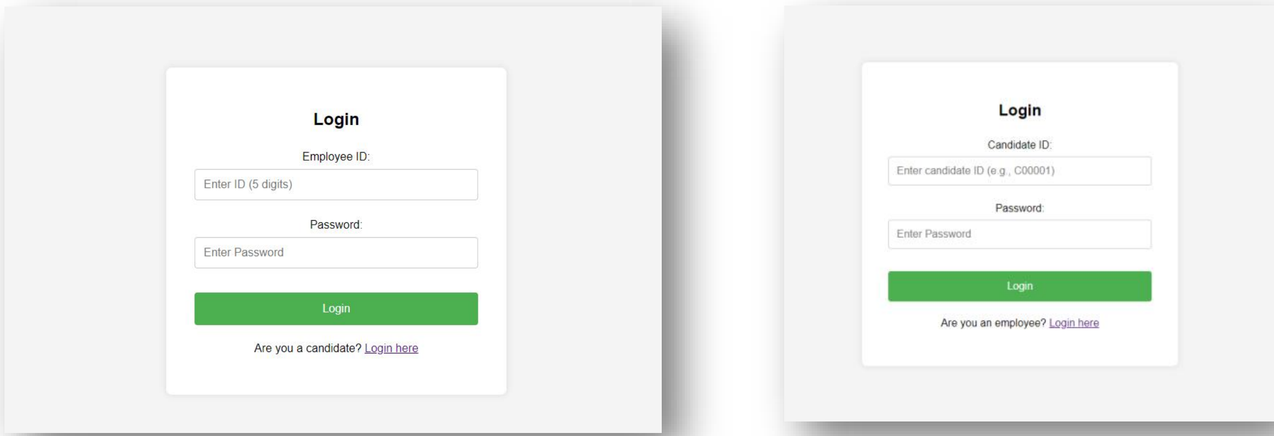


Fig 35. Login Page of Employee and Candidate

II. Admin Dashboard and Tasks:

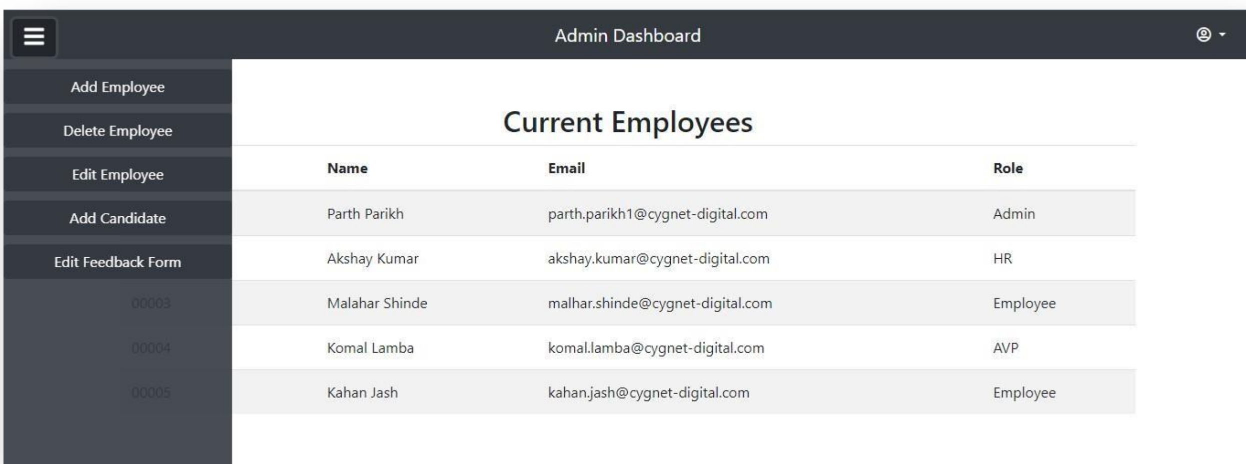


Fig 36. Admin Dashboard

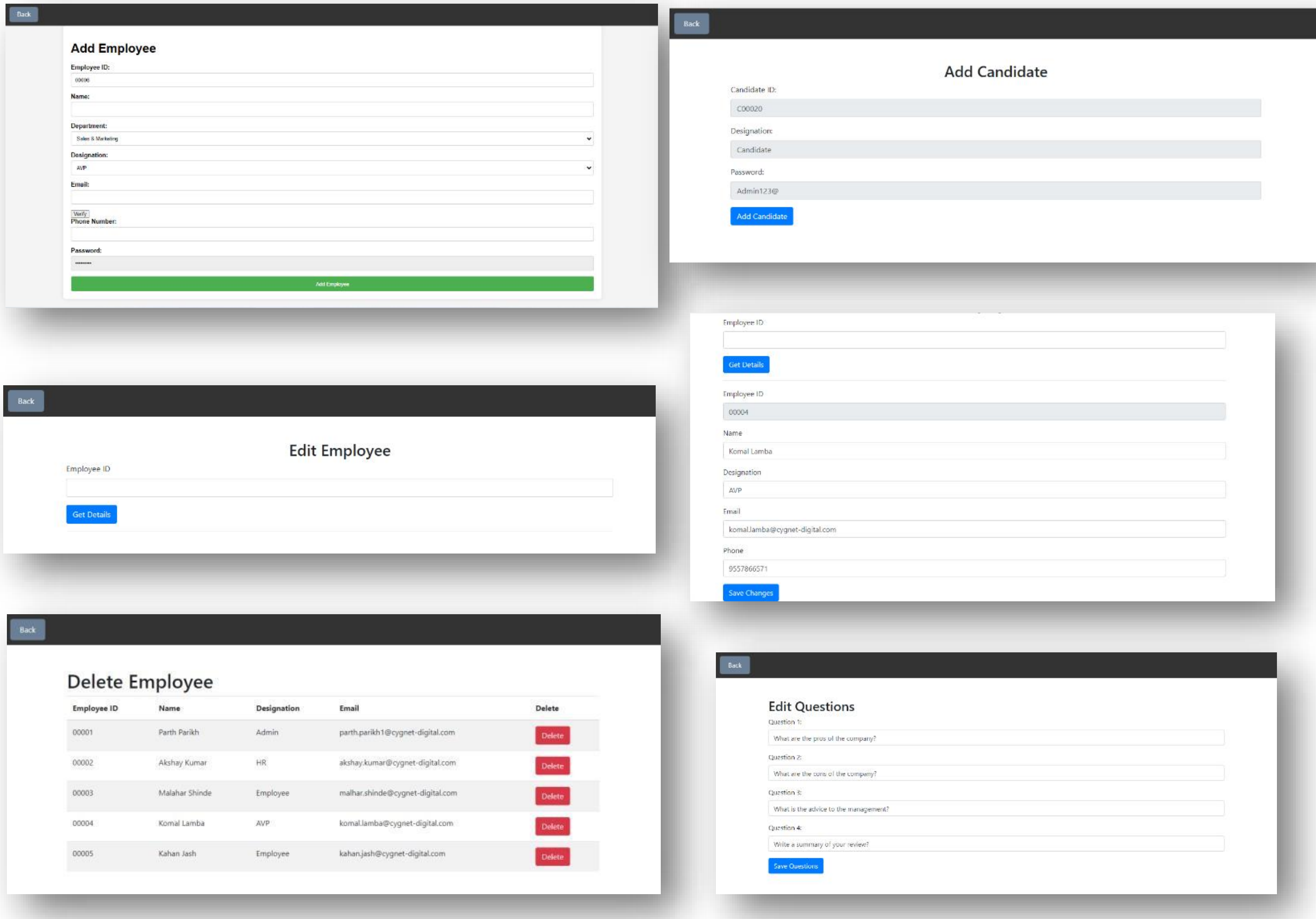


Fig 37. Admin Functionalities

III. Candidate Dashboard:

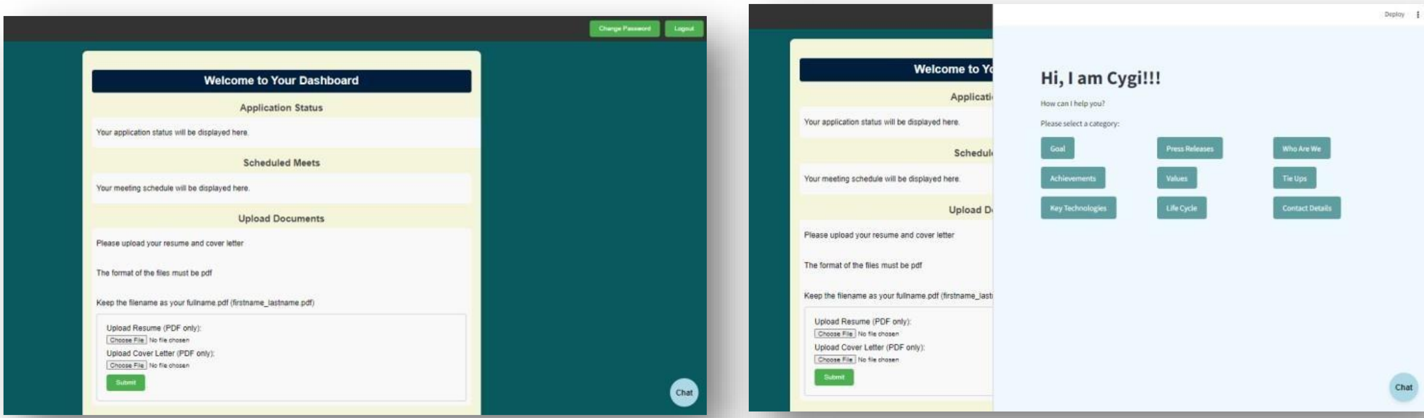


Fig 38. Candidate Dashboard

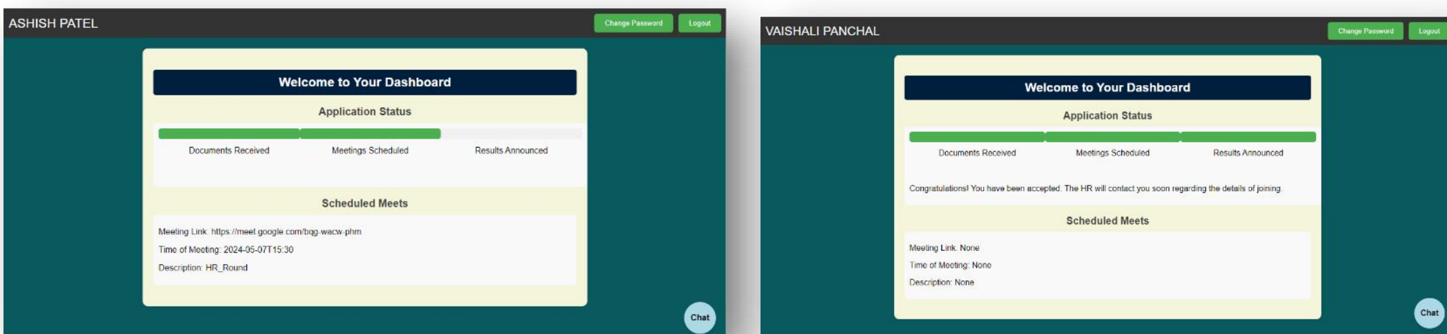


Fig 39. Glimpse of a Candidate Who has Submitted the Documents

IV. Employee Dashboard:

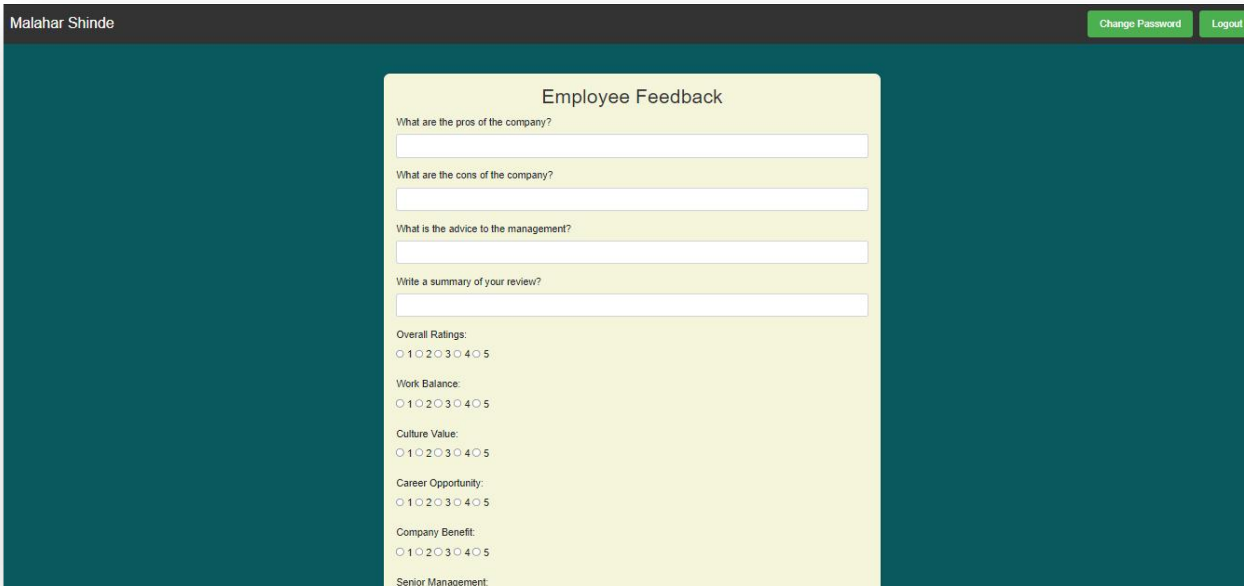


Fig 40. Employee Dashboard

V. HR Dashboard and Tasks:

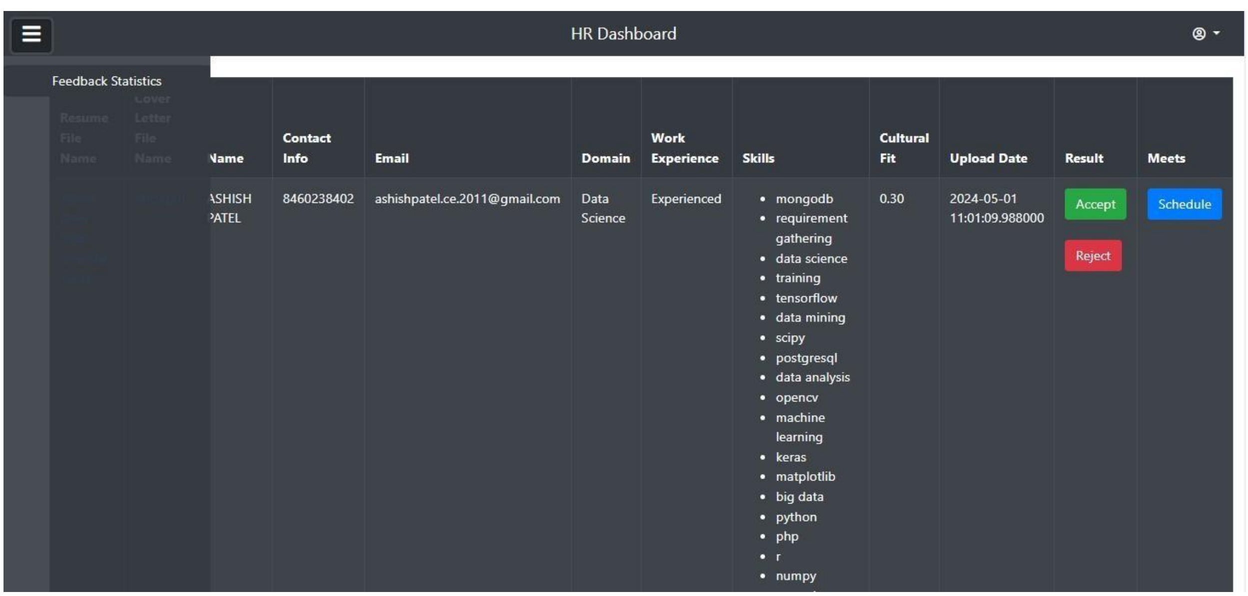


Fig 41. HR Dashboard

Fig 42. Meeting Scheduling Page

On the meeting scheduling page, the HR has to enter the details of the meeting including the meeting Link, date and time of the meeting and a brief description of the meeting.

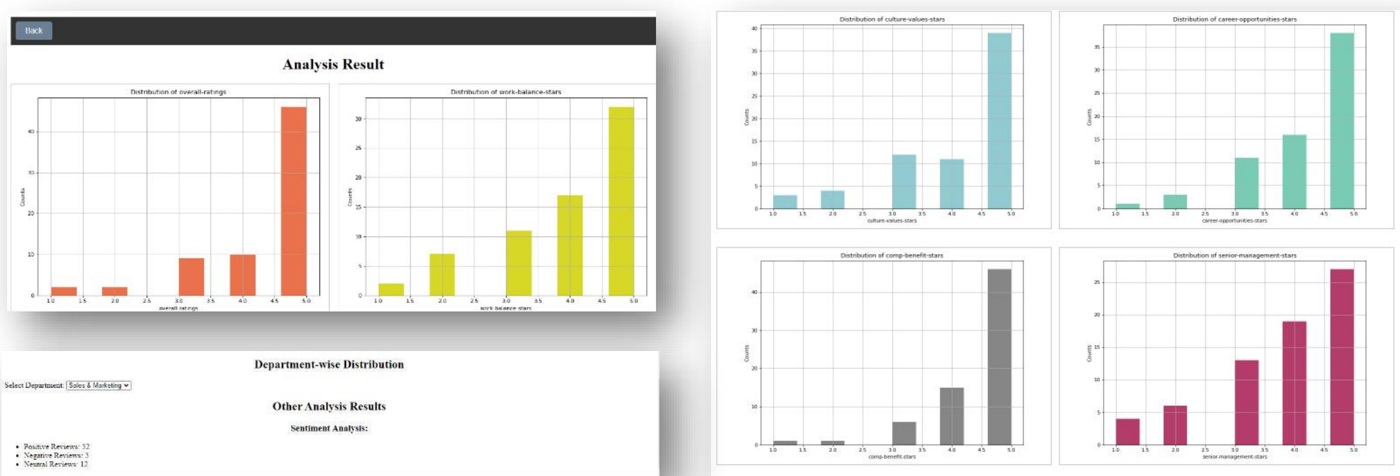


Fig 43. Employee Reviews Analysis

VI. AVP Dashboard and Tasks:

The AVP Dashboard features promotion prediction predicted by the promotion prediction module, displaying a roster of employees eligible for promotion based on uploaded CSV data encompassing metrics such as KPIs met, awards received, average training attendance, training scores, tenure, and age. The AVP retains the authority to accept or reject promotions, with the table listing employee IDs alongside department names. This structured approach ensures efficient management of promotion decisions, leveraging comprehensive data analysis to drive strategic personnel advancements within the organization.

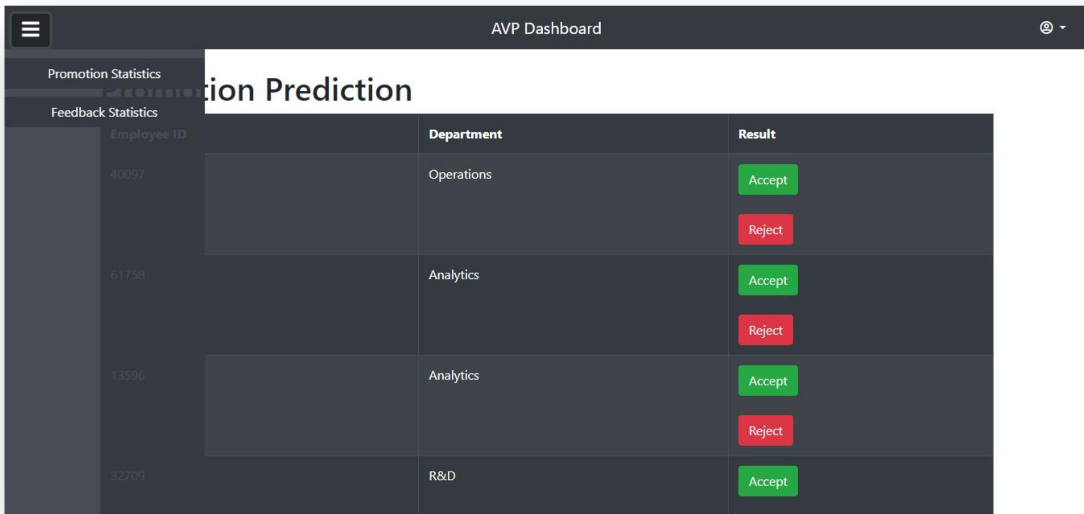


Fig 44. AVP Dashboard

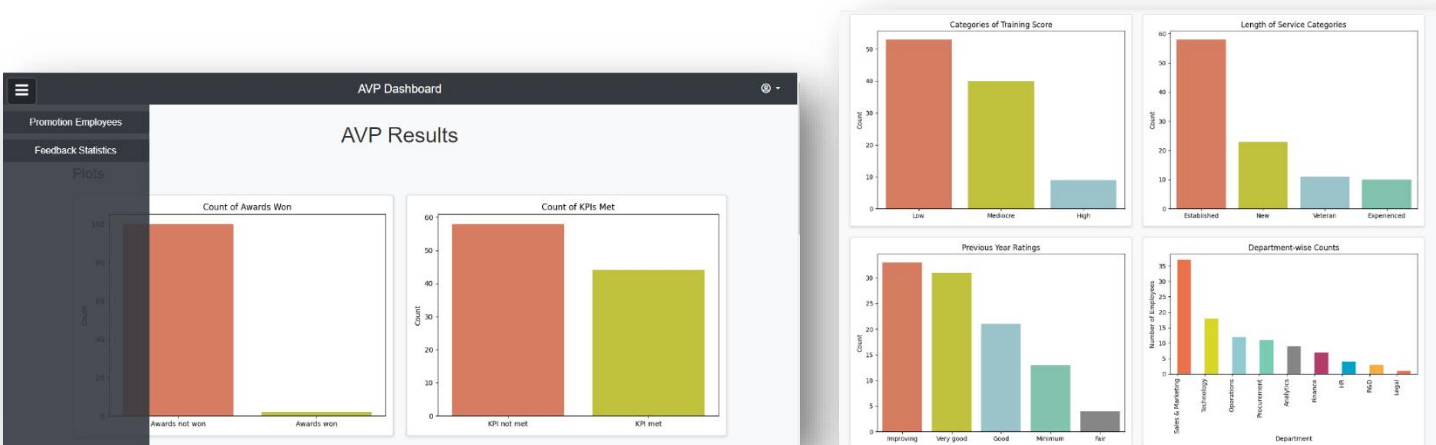


Fig 45. Employee Promotion Analysis

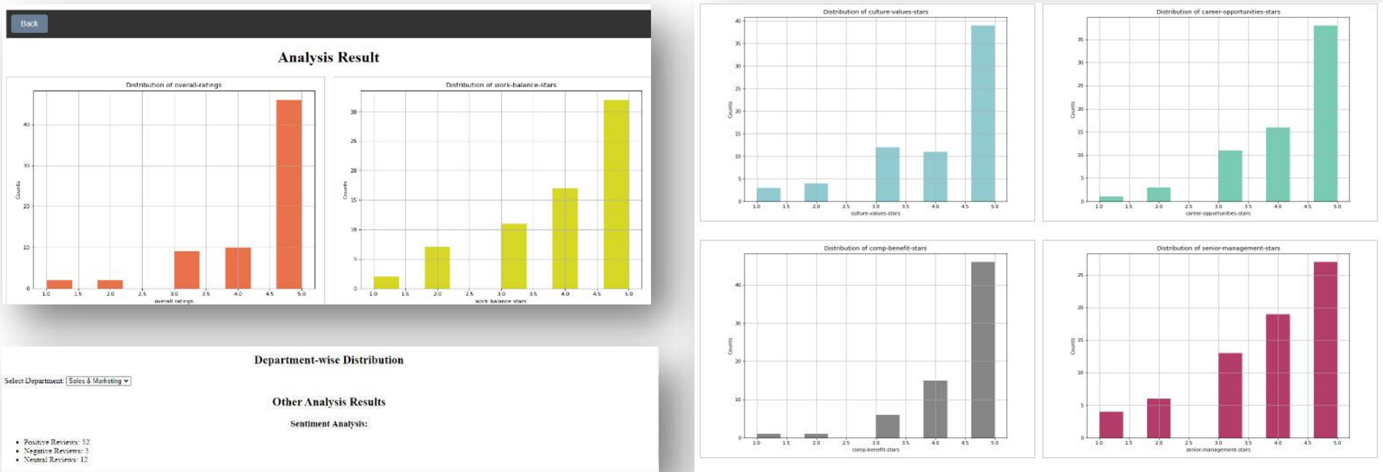


Fig 46. Employee Feedback Analysis

VII. Change Password Page:

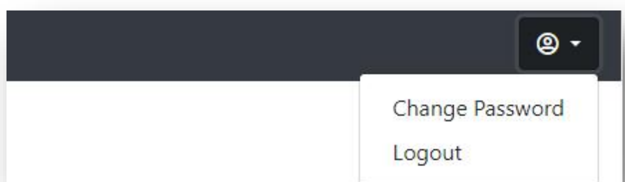


Fig 47. Change Password Button

In every module, including candidate, admin, AVP, and HR, the top bar's rightmost corner displays buttons for changing password and logging out.

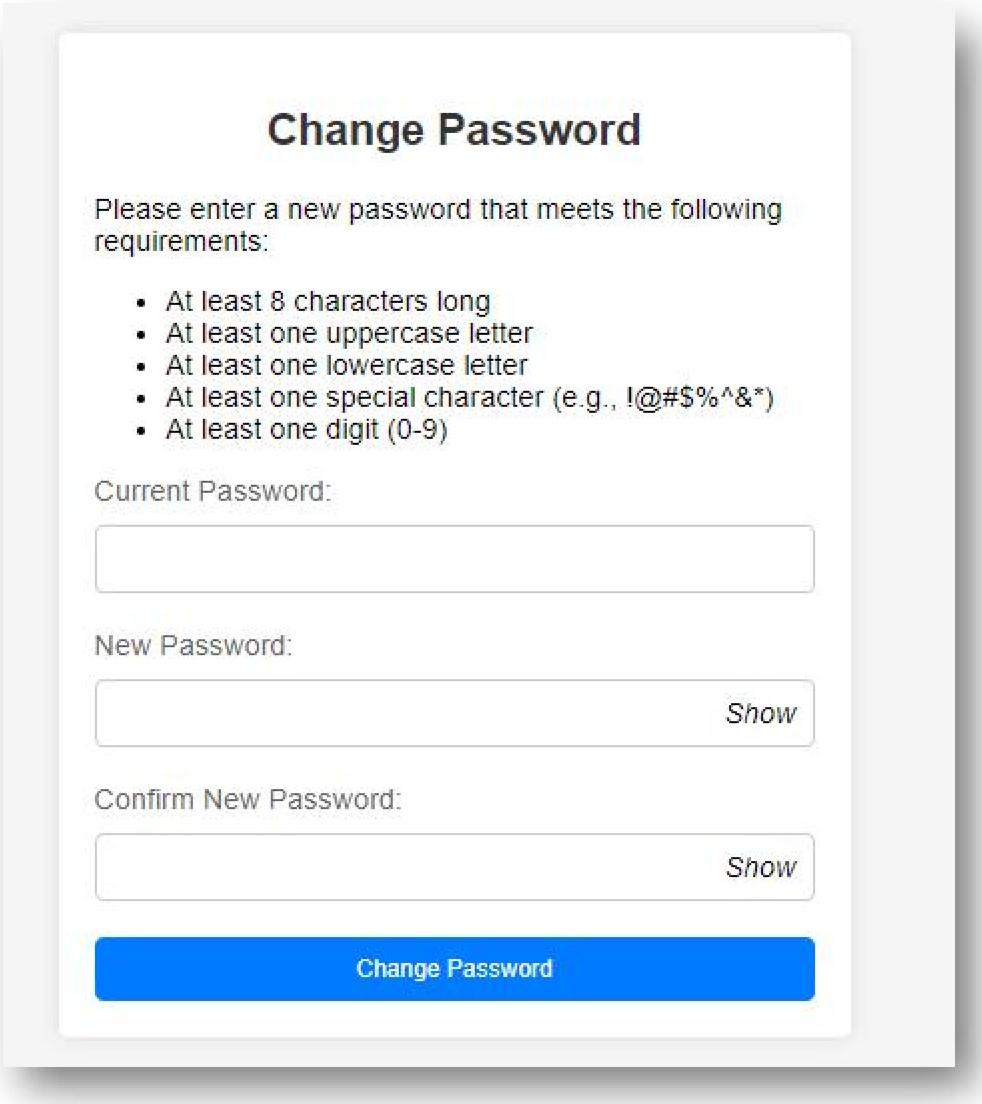


Fig 48. Change Password Page

Clicking the "Change Password" button in any module leads to a dedicated page where users can modify their passwords, adhering to specified criteria. These include a minimum length of 8 characters, inclusion of at least one uppercase letter, one lowercase letter, one special character, and one digit. Failure to meet these requirements prevents password alteration. Following modification, users must confirm the new password to effectuate the change successfully. This structured process ensures password security and adherence to robust authentication standards across all modules within the system.

Chapter 5
Conclusion and Future Scope

5.1 Conclusion

In the realm of modern human resources, the integration of AI-driven HR analytics stands as a transformative force, reshaping conventional practices and ushering in a new era of workforce management. Leveraging sophisticated technologies such as natural language processing, sentiment analysis, and predictive analytics, organizations can streamline and optimize various HR processes.

Through the automation of tasks like resume screening and interview scheduling, considerable time and resources are saved, while simultaneously ensuring fairness and objectivity in candidate selection. This not only accelerates the recruitment process but also enhances the overall candidate experience, setting a positive tone for their potential engagement with the organization.

Additionally, the ability to analyze employee sentiment and predict turnover risks empowers HR professionals to proactively address retention challenges, thereby fostering a more engaged and loyal workforce. By identifying potential issues before they escalate, organizations can implement targeted interventions and initiatives to improve employee satisfaction and reduce turnover rates.

Moreover, the real-time data visualization capabilities provided by AI-enabled HR analytics offer invaluable insights into workforce dynamics and trends. Through interactive dashboards and visual representations, HR professionals gain a comprehensive understanding of key metrics, allowing for informed decision-making and strategic planning.

Despite challenges related to data quality and integration, the benefits of AI in HR analytics are undeniable. By embracing these advanced technologies, organizations gain a competitive edge in talent acquisition, retention, and management.

As businesses navigate an increasingly complex and dynamic landscape, the adoption of AI-driven HR analytics emerges as a critical imperative for driving sustainable growth and success. By harnessing the power of AI, organizations can unlock the full potential of their human capital, fueling innovation, productivity, and organizational excellence in the ever-evolving digital age.

5.2 Future Scope

The future scope for the AI-enabled HR analytics project is promising and multifaceted. Firstly, continued advancements in AI and machine learning algorithms offer opportunities to enhance the accuracy and efficiency of HR processes further. Refinement of natural language processing models can enable deeper insights from candidate resumes and cover letters, while advancements in predictive analytics can facilitate more precise identification of retention risks and talent development opportunities.

Furthermore, the integration of emerging technologies such as computer vision and sentiment analysis from multimedia sources presents avenues for richer candidate assessments and employee feedback analysis. By incorporating visual data from video interviews and social media platforms, HR analytics can gain deeper insights into candidate personalities, cultural fit, and employee sentiments, thereby enriching decision-making processes.

Moreover, the project's future scope extends beyond recruitment and retention to encompass broader HR functions such as performance management, learning and development, and workforce planning. AI-driven tools can support personalized learning pathways, identify skill gaps, and predict future workforce needs based on evolving business requirements and industry trends.

Additionally, there is potential for collaboration with other organizational departments, such as finance and operations, to leverage HR analytics for cross-functional insights and strategic alignment. By integrating HR data with key performance indicators and business metrics, organizations can optimize resource allocation, mitigate risks, and drive holistic organizational performance.

Lastly, ongoing advancements in data privacy and ethics frameworks will be critical to address concerns surrounding algorithmic bias, data security, and compliance with regulatory requirements. By prioritizing transparency, accountability, and ethical AI practices, organizations can foster trust and confidence in the use of AI-enabled HR analytics, paving the way for sustainable growth and innovation in the future.

References:

[1] https://www.anaconda.com/download/

[2] https://www.coursera.org/specializations/natural-language-processing

[3] (24) Live NLP Playlist - YouTube

[4] Convolutional Neural Networks | Coursera

[5] Sequence Models | Coursera

[6] What is Scrum? [+ How to Start] | Atlassian

[7] What is HR Analytics? | Definition, Examples, Metrics (valamis.com)

[8] AI in HR Analytics: Benefits, Challenges & Ways to Integrate (pockethrms.com)

[9] MongoDB Tutorial (w3schools.com)

[10] https://www.kaggle.com/datasets/gauravduttakiit/resume-dataset

[11] ShashiVish/cover-letter-dataset · Datasets at Hugging Face

[12] Pradeep Kumar Roya, Sarabjeet Singh Chowdharyb, Rocky Bhatiab- A Machine Learning approach for automation of Resume Recommendation system

[13] Python RegEx (w3schools.com)

[14] https://www.pragnakalp.com/case-study/nlp-resume-parser-bert-python/

[15] https://medium.com/@2020.chetaniya.bajaj/building-a-resume-parser-using-nlp-dd36c2afbce9

[16] https://huggingface.co/sentence-transformers

[17] https://www.geeksforgeeks.org/cosine-similarity/

[18] We are Cygnet, We are One

[19] Sangita Pokhrel , Swathi Ganesan , Tasnim Akther , Lakmali Karunarathne - Building Customized Chatbots for Document Summarization and Question Answering using Large Language Models using a Framework with OpenAI, Lang chain, and Streamlit

[20] (24) LangChain - YouTube

[21] Generative AI with Large Language Models | Coursera

[22] Employees-Reviews dataset (kaggle.com)

[23] Babacar Gaye, Dezheng Zhang, and Aziguli Wulamu - Sentiment classification for employees reviews using regression vector- stochastic gradient descent classifier (RV-SGDC)

[24] https://towardsdatascience.com/predicting-sentiment-of-employee-reviews-ec0c0c837328

[25] raw.githubusercontent.com/rajtulluri/Employee-Promotion Prediction/master/employeePromotion.csv

[26] (24) Flask - YouTube

[27] HTML Tutorial (w3schools.com)

[28] https://www.w3schools.com/css/default.asp