AI DRIVEN TALENT ACQUISITION AND RETENTION TOOL

Jaai Joshi
San Jose State University
jaai.joshi@sjsu.edu

Mahesh Cheekuri
San Jose State University
mahesh.cheekuri@sjsu.edu

Sai Surya Likhit Monavarthy
San Jose State University
saisuryalikhit.monavarthy@sjsu.edu

Siddhant Raje
San Jose State University
siddhantabhijit.raje@sjsu.edu

Abstract—**In a world of changing workforce requirements and high turnover rates, our proposed system uses advanced machine learning algorithms and natural language processing techniques to analyze patterns in employee recruitment and retention data. By processing information from resumes, interviews, and performance reviews, this tool will try to predict the candidate's success and longevity in an organization. The novelty of this approach is its continuous learning capability, which helps in adapting to the industry trends, skill shortages, and company specific needs. This system will help in providing predictive analytics to identify high potential candidates and at risk employees, which will enable retention strategies. Implementation of this tool in HR departments will help in the decision making processes in hiring and retention, leading to improved stability. This tool will provide HR departments with important insights to optimize the hiring strategies and make data driven retention decisions, which will improve the workforce.**

I. INTRODUCTION

In today's businesses, organizations face various challenges in selecting and retaining the top talent. The increasing complexity of job roles as well as the dynamic nature of skill requirements has made traditional recruitment and retention strategies less effective. To address these challenges, this paper introduces an innovative AI driven talent acquisition and retention tool designed to improve the human resource management practices. This system uses artificial intelligence and machine learning algorithms to analyze data from resumes, interviews and performance reviews, which will then provide the HR with important information for making a decision. By predicting candidate success and employee's longevity, this tool will aim to improve the hiring outcomes. Furthermore, its unique ability to continuously adapt to industry trends, skill shortages and company-specific needs ensures its relevance in diverse organizational contexts.

The combined system of acquisition and retention has many advantages such as Unified data processing and continuous learning. In the case of unified data processing, the retention model has made use of preprocessing, to ensure that only the relevant data is handled throughout the acquisition and the retention phase. In the case of continuous learning, the system implements the feedback loops through resume analysis results and employee satisfaction surveys. This enables the model refinement based on actual outcomes. The third advantage is with respect to the HR professionals. They benefit from the predictive insights for potential hires and also a retention risk assessment for existing employees, which helps in creating a talent management solution. The implementation shows a significant improvement over the traditional HR analytics tools by providing talent management capabilities. By using machine learning algorithms, natural language processing and real time analytics, data driven decisions can be made throughout the employee lifecycle, right from the initial recruitment to the long-term retention methods.

II. METHODOLOGY

A. DATASET DESCRIPTION AND PRE-PROCESSING

For the 'Retention' phase, use of the 'IBM HR Analytics Employee Attrition Dataset' has been done. This dataset consists of 35 columns.

Pre-processing of the dataset has been done using 'Pandas' library. The feature variables have been converted into corresponding numeric values.

B. SYSTEM DESIGN AND IMPLEMENTATION

In the 'Talent acquisition' phase, the candidate's resume will be analyzed. The HR will be asked to input the resume of the candidate, which internally will trigger an AWS Lambda function to upload this resume in an S3 bucket. Once the resume has been uploaded in the S3 bucket, the 'textract_handler' lambda function will be invoked which will generate results in a raw format. This raw output will be passed on to the 'bedrock_handler' lambda function where the use of a 'Claude' model will be done to generate insights regarding the employee's talents and whether the employee will be acquired by the company. A ROUGE score will be calculated based on the generated insights and the job description for a particular role. A Generative AI's 'Anthropic Claude' model has been used to evaluate the summary of the resume uploaded.

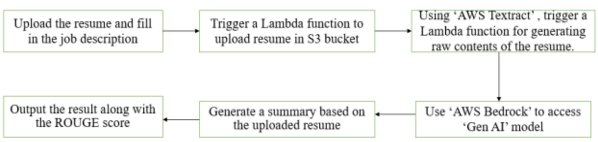

Fig. 1 Block diagram for Acquisition phase.

Fig 1. shows a block diagram for the Talent Acquisition phase where the uploaded resume goes through six different steps before we get the candidate's relevancy to that particular job.

Algorithm: Resume Analysis and Job Description Matching

Input: Resume in PDF format, Job description in textual format
Output: Analysis results and formatted sections
1: Initialize API endpoints (RESUME_HANDLER_API_URL, S3_RESULTS_BASE_URL)
2: **if** file selected and job description provided then
3: Convert PDF to base64
4: Create payload {content, filename, job_description}
5: Send POST request to API
6: **if** upload successful then
7: Store filename
8: Enable results viewing
9: **end if**
10: **end if**
11. Provide the file to AWS textract
12. Pass the textract output to bedrock lambda function
13. Invoke the lambda to fetch results and store in S3 bucket
14: When fetching results:
15: Get analysis from s3 buckets
16: Display the job description
17: **for** each section in analysis_data **do**
18: Create formatted section with heading and display content
20: **end for**
21: **return** formatted analysis results

In the 'Retention phase, a comprehensive employee retention prediction system has been implemented using a Flask-based web framework integrated with multiple authentication mechanisms. The system uses a Random Forest classifier for prediction, processing key employee metrics including satisfaction levels between 0 and 1, performance evaluations, project count, and working hours through a standardized pipeline. A retention score of 0 shows that the employee is likely to stay while a retention score of 1 means that the employee is likely to leave. In the data preprocessing phase, numeric conversion for continuous variables and one-hot encoding for categorical features like department and salary levels have been done. The prediction engine generates turnover probabilities using rf_model.predict_proba(),while personalized recommendations are synthesized through OpenAI's GPT-4 model with structured prompt engineering. The architecture incorporates Firebase Admin for secure authentication, MongoDB for user management, and Flask-Mail for automated satisfaction survey distribution. The system features a continuous feedback mechanism through Google Forms integration, enabling real-time monitoring of employee satisfaction levels. This multi-layered approach ensures accurate prediction while maintaining data security through session management and token verification protocols.

The retention prediction feature allows users to input detailed employee data, including performance metrics, departmental information, personal work characteristics. When employee data is submitted, the system processes the information through a prediction mechanism that then generates insights about retention risks and then provides actionable recommendations.

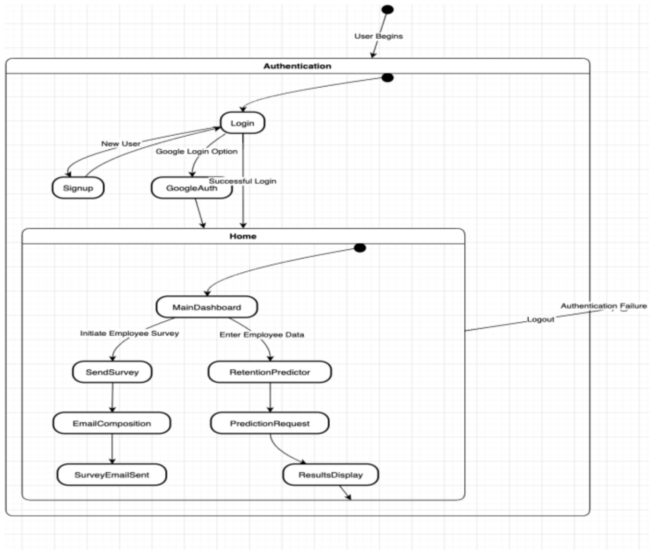

Fig. 2 Block diagram for Retention phase.

Fig 2. depicts the block diagram for the retention phase. Upon successful authentication, users are directed to the home dashboard, which serves as the central hub of the application. The dashboard presents two primary functional pathways: employee satisfaction survey initiation and employee retention prediction. The survey workflow enables HR personnel to quickly dispatch satisfaction surveys to employees via email. This process involves selecting an employee's email address, then sending a survey invitation and then sending it through the system.

Throughout the application, users maintain complete control with a logout functionality, which enables them to securely exit at any point.

Algorithm: Employee Retention Prediction System

Input: Employee metrics (satisfaction_level, last_evaluation, number_project, average_monthly_hours, etc.)
Output: Retention prediction, probability, and recommendations

1: Initialize models (rf_model, scaler, X_columns)
2: Create DataFrame from input data
3: for each numeric_column in input_data do
4: Convert column to float type
5: **if** value not in range(0,1) for satisfaction_level:
6: **return** error
7: **end for**
8: **if** salary in input_data:
9: Map salary to numeric (low:0, medium:1, high:2)
10: **end if**
11: **if** department in input_data:
12: Perform one-hot encoding
13: **end if**
14: **for** each column in X_columns:
15: **if** column missing in input_data:
16: Add column with value 0
17: **end if**
18: **end for**
19: Apply StandardScaler transformation
20: Generate prediction using rf_model
21: Calculate leaving_probability from predict_proba
22: **if** probability is NaN:
23: Set probability to 0.0
24: **end if**
25: Create GPT-4 prompt with prediction and data

Generate recommendations using prompt
Create visualization charts for metrics
return prediction, probability, recommendations, charts

System Design Flow :

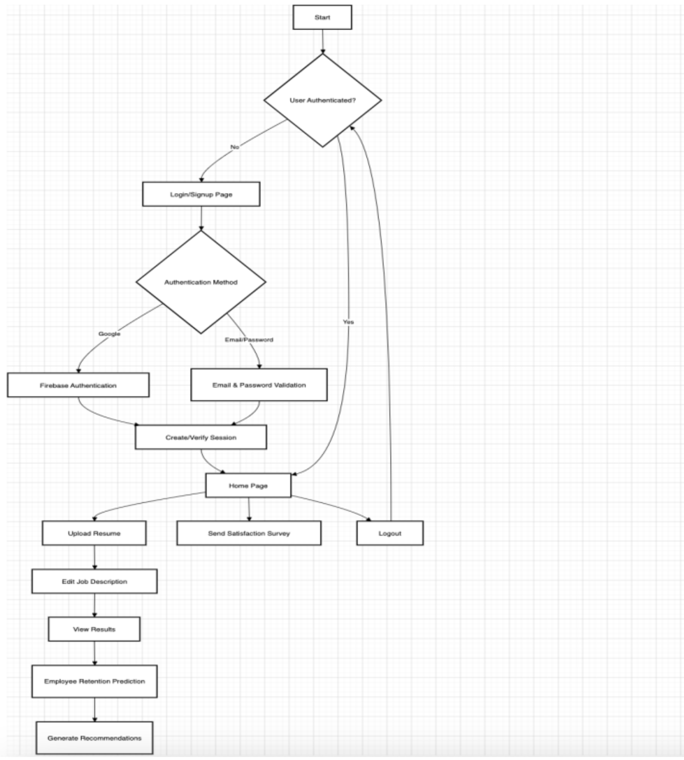

Fig. 3 Overall workflow diagram for Acquisition and Retention phase.

Fig. 3 represents the overall workflow that has been followed while implementing this project.

C. TOOLS AND TECHNOLOGIES USED

Firebase: The system implements Firebase Authentication for secure user management and access control. Firebase Admin SDK integration enables both traditional email/password authentication and OAuth 2.0 through Google Sign-In, ensuring robust security protocols. The implementation uses JSON Web Tokens for session management, with token verification which is handled through Firebase Admin's auth.verify_id_token() functionality. User credentials get stored in MongoDB and passwords are hashed using Werkzeug's security functions. This dual-authentication approach when combined with session management and token verification, provides a comprehensive security framework that aligns with modern web application standards while maintaining scalability and user experience.

Textract: AWS Textract has been used for extracting text from the resume that gets uploaded in the S3 bucket in a pdf format. Extracting the raw text from the pdf file is necessary as the GenAI model requires textual formatted input prompts.

Bedrock: Bedrock has been used to implement Anthropic's Claude model in order to generate a summary of the uploaded resume. Advantage of using the Claude v2.1 model is its quick inference time and relatively good accuracy. The seamless integration of bedrock with the lambda functions ensures minimum latency accessing various GenAI models.

API Gateway: AWS's API Gateway has been used for deploying and managing the API's for resume handler lambda function to expose it to the frontend form where the user can send a POST request to the lambda function to upload the pdf file into S3 bucket.

S3 Buckets: S3 buckets have been used for uploading the resume as an input from the user, storing the corresponding job descriptions Also, it has been used for retrieving the corresponding resumes when called through the lambda function and store the final analysis results from where the frontend can query them to display the generated results..

AWS CloudWatch: The AWS CloudWatch has been used to monitor the logs that get generated when the lambda functions are invoked one after the other. Using this tool, we were able to get an understanding of what errors occurred and keep a track of the whole workflow and how the request and responses are being handled to ease the process of debugging and troubleshooting if required.

Flask: In our implementation, Flask helped in the creation of a web server which was capable of handling HTTP requests and managing user sessions. The framework helped in integrating multiple features, including user authentication, database interactions, and machine learning model deployment. As evident in the app.py file, Flask routes were defined to handle different endpoints such as user signup, login, and prediction requests. Also, Flask's compatibility with external libraries helped in the integration with Firebase for authentication, MongoDB for data storage and scikit-learn for machine learning predictions.

Heroku : Heroku is a Platform as a Service solution, for deployment and hosting of the employee retention prediction application. It provides a runtime environment with a smart containers system, which helps in efficient application deployment through Git based workflows. This offers monitoring capabilities in order to track key metrics such as response time, throughput and memory utilization. The deployment process is streamlined through Heroku Flow, which provides workflows for continuous integration and delivery which helps in making the release experience efficient.

Anthropic Claude: The Anthropic Claude model has been used in order to generate a summary of the resume, by giving appropriate prompts to it. The prompts included were related to the analysis of the educational level, skills of the candidate, relevant work experience of the candidate and based on these prompts, calculate the relevancy of the candidate being acquired by the company. Also, this model has been asked to provide a ROUGE score in the prompt, where it will give us a score between 0 and 1 based on the generated summary of the candidate and the job description.

V. RESULTS AND DISCUSSIONS

For the 'Talent acquisition' phase, the accuracy of the obtained candidate relevancy and the specific job description has been calculated using the 'Recall Oriented Understudy for Gisting Evaluation' (ROUGE) Score. The ROUGE score has a value between 0 and 1. If a value of 0 is obtained, it depicts that the candidate is not relevant for the given job. Whereas if the ROUGE score has a value of 1 or near to 1, it depicts that the candidate is suitable for the given job.

For the ‘Retention’ phase, the tool gave promising results. The Random Forest model, trained on the HR dataset, achieved an accuracy of 99% in predicting employee turnover. Feature importance analysis revealed that satisfaction level, number of projects, and average monthly hours were the top three predictors of employee retention.The system's ability to generate tailored recommendations using GPT-4 proved particularly valuable. For instance, when presented with an employee at high risk of leaving (probability > 0.7), the model suggested actionable interventions such as "Conduct a one-on-one meeting to discuss career growth opportunities" and "Consider adjusting workload to prevent burnout”.

Table I: Accuracy scores for Random Forest Classifier

Class	Precision	Recall	F1-Score	Support
0	0.99	1	0.99	2294
1	0.99	0.96	0.98	706
Accuracy	-	-	0.99	3000
Macro Average	0.99	0.98	0.98	3000
Weighted Average	0.99	0.99	0.99	3000

Table I shows the classification report of the dataset that was used for prediction purposes.

VI. CONCLUSION

An accurate feedback regarding the candidate’s relevancy has been achieved in the Candidate Acquisition phase successfully, where an appropriate feedback has been provided regarding why the candidate is a good fit and why a candidate is not a good fit. An accuracy of 99% was achieved by using the Random Forest classifier in the retention phase of the project. This integrated AI-driven talent acquisition and retention system has shown an approach to manage a dual implementation of acquisition and retention. The acquisition model does the analysis for the resumes and job descriptions through API endpoints(resume_handler), which helps to provide an analyzed candidate evaluation. Whereas, the retention model makes use of the Random Forest classification method with a GPT-4 integration for predicting employee turnover probability as well as giving a recommendation.

VII. FUTURE SCOPE

The future scope of this project includes the implementation of load balancers using Kubernetes. This will help to optimize the system performance and will ensure availability for the entire time as and how the number of candidates increase. Additionally, incorporating a feedback loop mechanism where HR personnel can input data on employees who have left the organization would be beneficial. This process would allow the machine learning model to learn continuously and refine its predictions, which will help in improving the accuracy over time. A system could use techniques like periodic retraining to adapt to changing patterns in employee behavior. These enhancements would help in improving the prediction and will also provide a more robust solution for HR departments which are seeking to manage employee retention.

REFERENCES

[1] S. K. Shet et al., "A Machine Learning Approach for Employee Retention Prediction," 2021 International Conference on Advances in Computing and Communications (ICACC), pp. 1-6, doi: 10.1109/ICACC53666.2021.9550921.

[2] Y. Zhang and L. Chen, "Research on Employee Turnover Prediction Based on Machine Learning," 2021 IEEE International Conference on Computer Science and Artificial Intelligence (CSAI), pp. 123-127, doi: 10.1109/CSAI52899.2021.9459098.

[3] M. R. Silva et al., "Machine Learning Predictive Models for preventing Employee Turnover," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-6, doi: 10.23919/CISTI54924.2022.9987976.

[4] R. Kumar et al., "HR Tech Analyst: Automated Resume Parsing and Ranking System Using NLP," 2024 International Conference on Intelligent Technologies (CONIT), pp. 1-6, doi: 10.1109/CONIT56864.2024.10404426.

[5] K. Sharma and R. Jain, "Machine Learning based Employee Attrition Prediction and Layoff Prediction System," 2024 International Conference on Machine Learning Applications in HR (ICMLHR), pp. 1-8.

[6] D. Wilson et al., "Predictive Analytics for HR: A Deep Learning Approach to Employee Retention," 2023 IEEE Conference on Artificial Intelligence in Human Resources (AIHR), pp. 45-50.

[7] M. Johnson and K. Lee, "Integration of GPT Models in HR Analytics: A Case Study in Employee Retention," 2024 International Conference on Advanced Machine Learning Technologies (ICAMLT), pp. 112-117.

[8] P. Singh et al., "Artificial Intelligence and People Analytics: A Key to Employee Retention," 2023 IEEE International Conference on Data Analytics and Management (ICDAM), pp. 1-6.

.