BSCS-F24-020

03-134212-066 Muhammad Mughees Ahmed

03-134212-067 Muhammad Owaim

03-134212-096 Wajeeha Khan

Automated HR

In partial fulfilment of the requirements for the degree of

Bachelor of Science in Computer Science

Supervisor: Muhammad Mudassar

Department of Computer Sciences

Bahria University, Lahore Campus

June 2025

Certificate

We accept the work contained in the report titled

“Automated HR”

written by

Muhammad Mughees Ahmed

Muhammad Owaim Wajeeha

Khan

as a confirmation to the required standard for the partial fulfilment of the degree of

Bachelor of Science in Computer Science

Approved by:

Supervisor: Muhammad Mudassar

(Signature)

June, 2025

DECLARATION

We hereby declare that this project report is based on our original work except for citations and quotations which have been duly acknowledged. We also declare that it has not been previously and concurrently submitted for any other degree or award at Bahria University or other institutions.

Enrolment	Name	Signature
03-134212-066	Muhammad Mughees Ahmed	
03-134212-067	Muhammad Owaim	
03-134212-096	Wajeeha Khan	

Date : June 05, 2025

Especially dedicated to my beloved

grandmother, mother and father

(Muhammad Mughees Ahmed)

my beloved grandmother, mother and father

(Muhammad Owaim)

my beloved grandmother, mother and father

(Wajeeha Khan)

ACKNOWLEDGEMENTS

We would like to thank everyone who contributed to the successful completion of this project. We would like to express our gratitude to my research supervisor, Mr. Muhammad Mudassar for his invaluable advice, guidance and his enormous patience throughout the development of the research.

In addition, we would also like to express my gratitude to our loving parents and friends who had helped and given me encouragement.

Muhammad Mughees Ahmed
Muhammad Owaim
Wajeeha Khan

Automated HR

ABSTRACT

The Automated HR System is a significant step forward in how organizations handle their recruitment processes, without the bulky manual workflows, and replacing the same with a fully intelligent platform that makes extensive use of data. Essentially, the system ingests both resumes and job descriptions, followed by the use of advanced NLP methods to extract important data out of unstructured text – for instance capabilities, qualifications, or work history. This information is passed through sophisticated machine learning models such as XGB that have been trained to identify patterns and prediction of a candidate-job match using history hiring data. Consequently, each applicant receives a robust relevance score that implies not only the fit between his/her background and the requirements of the position, but also finer hints like career development or compatibility of soft skills derived from the content of the resume.

After the evaluation of the candidates, the platform automates the entire shortlisting and notification workflow. High scoring applicants are marked for the hiring managers right away with those who fail the marking threshold receiving polite, personalized updates – making sure that no candidate is left without timely feedback. Automated communication obviates redundant administrative operations and greatly decreases the likelihood of error or failure to notice mistakes that result from human error. In addition, since everything is approved by measurable criteria, organizations can be sure their selection process is not only objective but also identical for all, thus curtailing unconscious biases all too present in manual reviews.

Apart from the short-term benefits in speed and accuracy, Automated HR System has scalability and long-term strategic vision. Detailed dashboards summaries performance over several recruitment cycles, so that HR leaders can look at metrics like time to fill, candidate quality, and conversion rates. These modules of predictive analytics use this historical data to make predictive analysis on the hiring demands and future skill shortage to make companies plan workforce strategies preemptively, and not reactively. Coming updates though are promising even more integration with enterprise level human capital management software, uniting recruitment, measurement of

performance, workforce planning, into a single ecosystem. Basically, what the automated HR System does is that it makes traditional hiring a dynamic and a learning process. Utilizing the power of natural language processing, machine learning and real time analytics, it speeds up time-to-hire, reduces costs and raises quality of selection of candidate selection. Meanwhile, it gives the HR teams the strategic insights that they need for anticipating future talent requirements and creating a more agile and resilient workforce.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Background

The Human resources wing of any organization has a double role to play. attraction of best talent by having a good look at high quality resumes and liaising with selected persons to offer them employment, as well as handling needs and concerns of current employees. A wholesome HR management system equips companies to be at the peak levels of performance because the company does not only hire the right target but also ensures that it develops clear blueprints of policies and procedures to ensure employee success. By contrast, old, manually - driven HR processes are crippled by repetitive, time-consuming processes which are prone to error, and inefficiency. Such period dated workflows tend to slow decision-making, impact with inconsistencies, adversely affect the effectiveness of the HR function. Automating routine tasks such as resume grading, candidate correspondence, and policy implementation, modern systems leave the HR professionals more space to focus on strategic initiatives, making the organization more agile, more accurate, and people-centric.

Organizations should be concerned with the question of assessing candidate resumes, that is, one of the key obstacles that HR organizations have to overcome. High number of job applications against each role results in the HR staff members having to resort to manual examination of resumes thus eating up time and creating biased behavioral elements in favor of candidate selection practices. Interview panels conduct random selection process without considering professional expertise that corresponds with the qualification of the candidates. Poor assessment of candidates in the course of hiring cycles translates to unfruitful hire decisions for filling up vacancies.

Employee onboarding is established as one of the most important basic issues the current workplace operation recognizes. New staffs needed someone to guide them about company rules and work environment and about their duties. AI-driven Chatbot's can automate all the FAQs, which take up valuable Human Resources team time. Automated procedures should be introduced to remove boring aspects of the work from human resource specialists so that they can progress into major HR leadership roles that aim at educating work forces and maintaining teams.

A system considered for Human Resources automates the application screening for jobs while taking care of candidate's correspondences and staff supervision through AI-driven integrated processes. The system shortens HR operational time with this efficient accurate selection mechanism of the eligible applicants which spurs concurrent workforce performance improvement.

1.2 Problem Statements

There are some issues that face conventional HR management with respect to determining the effectiveness of recruitment, precision in making decisions as well as overall productivity. Some of them include:

- **Inefficient Resume Screening:** HR departments tend to take a lot of time to manually go through a vast number of resumes for each job vacancy. This process is laborious, prone to exhaustion and human prejudice and may lead to the best candidates' rejection.
- **Lack of Automated Onboarding Support**: New recruits need to ask equally many redundant questions on company policies, work culture, and roles. HR specialists would otherwise have to deal with such queries manually, and this consumes time that could otherwise be used in making more strategic plans such as developing and retaining employees.
- **Manual and Error-Prone Workflows**: Critical HR operations like data input, shortlisting of the candidates, and onboarding of the employees are performed manually. Not only does this lead to delays and errors but it also has an administrative burden and reduces effectiveness of recruitment.

1.3 Aims and Objectives

The objectives are shown as following:

- **AI-Based Resume Screening**: Use Machine Learning (ML) algorithms to screen, sort and narrow the pool of candidates by job description or category, Skill and experience. This will do away with human screening and also save time in the processing and minimize the margin of error in a selection from an applicant pool.
- **AI Chabot for On boarding Support**: Use an AI-based Chabot that is NLP enabled to support new hires by answering FAQs, informing them of the

company policies, and helping them on-board based on the comments each person makes.

The proposed system comes with fewer responsibilities for the HR personnel as compared to the previous system, but it has a positive impact on the overall experience of the employees. Through optimizing recruitment efficiency and automation of major functions it reduces manual effort and makes decision making easy. Consequently, human resource management becomes more efficient and more accurate with the assistance of AI, so that organizations can make smarter hiring decisions with increased effectiveness of their operations.

1.4 Scope of Project

This AI & ML-powered HR platform starts with transforming recruitment process. using natural language processing, it processes and analyses every resume, extracts important information, like education, work history, certifications, then it uses supervised-learning classifiers to measure and priorities candidates with incredible precision. Instead of having to wade through the endless applications that come flooding in like an ocean, recruiters are now left to depend on dynamic algorithms that can prioritize and weight criteria, such as years of experience, certain forms of technical proficiencies, or background in leadership, and even group similar candidates together in order to find unknown pools of talent. This does not just cut the time-to-hire by up to half but also guarantees the best individuals end up at the top of the pile.

When candidates are onboarded, a smart chatbot will handle routine communications providing 24/7 support for answering questions on benefits enrollment, IT set up processes and compliance training schedule. Which is trained on the company's own policies and FAQs this conversational AI answers questions consistently and accurately on chat, email or integrated messaging platforms, therefore reducing the HR tickets by as much as 70%. It also onboards new hires into it through mandatory workflows – prompting new hires to submit required documents, complete training modules, and track their progress – while flagging any lags to HRs' attention to make sure that no one falls through the cracks.

The employees' management module in the background admins have access to is a refined one, making it easy to create, read on, update and delete personnel records. Staff groups can update individual profiles or can do bulk imports and exports, perfect for department reorganizations or new certification roll outs, without accessing a piece of code. All change made there is timestamped and logged with user id including, supporting audit requirements and maintaining data integrity and role based access controls ensure that sensitive information is not leaked to wrong users. When combined, these elements go on to change HR from a purely administrative function into a strategic business partner. Data-driven dashboards bring to the fore hiring bottlenecks, turnover trends, and skills gapes, equipping the leaders with the actionable insights for ahead of time workforce planning. Predictive analytics takes it further and projects attrition risk and draws attention to high-potential employees for selective development. Based on cloud-native architecture, the system is easy to scale, as headcount increases, and its modular AI services allow you to plug new functionality, such as sentiment analysis or performance-management recommendations, without ripping your system out and starting over. What you get is quicker hiring, a happier workforce, and more cognitively and agile HR processes, which provide genuine competitive advantage.

CHAPTER 2

LITERATURE REVIEW

2.1 Literature Review

The fuzzy-logic–based framework dynamically adapts to both external and internal organizational factors, offering enterprises a means to enhance human-resource utilization and overall managerial effectiveness, as evidenced by advanced modeling techniques [1]. Structural equation modeling of 274 IT professionals in Chennai shows that AI-enabled accuracy, automation, and real-time insights substantially contribute to time savings and cost reductions, whereas computing power and personalization exert a less direct effect, as indicated by the SEM analysis [5]. By integrating the Unified Theory of Acceptance and Use of Technology with risk-event chain graphs, a predictive model for identifying and mitigating risks in digital HR processes is achieved, as outlined through risk-event chain modeling [6]. Comparative evaluation of decision trees and ensemble methods reveals significant gains in efficiency, cost reduction, performance enhancement, and robust trend-detection capabilities, as substantiated through comparative algorithmic evaluation [7]. Leveraging support-vector machines, k-nearest neighbors, and Word2Vec

embedding's with cosine-similarity measures streamlines HR tasks, yielding time efficiency improvements and a classification accuracy of 91 %, as confirmed by performance metrics [8]. Combining named-entity recognition with TF-IDF and cosine similarity simplifies recruitment through automated information extraction and applicant ranking, as validated by extraction and similarity measures [9]. Early application of machine learning alongside spaCy's NER pipeline enhanced candidate matching and placement success, laying foundational groundwork for today's AI driven HR systems, as recognized in pioneering implementations [10].

2.2 Literature Table

Table 2. 1: A review of the relevant literature.

Study	Year	Model	Findings	Results	Dataset
P. Yang et al. [1]	2024	Fuzzy Logic	It adjusts to both internal and external factors affecting personnel management, which are considered critical for development of human resource practices in the contemporary enterprises. The system also opens the window for enhancing the effectiveness in utilization of human resources in the organization.	N/A	Private
Nawaz et al. [2]	2024	SEM method	Integration of AI in the HRM enhances accuracy, automation, and real-time responsiveness- which saves a lot of time and money- whereas computing power and personalization play a smaller role. An obtained model with the use of 274 IT professionals in Chennai confirms that targeted automation does ravel HR processes as well as bolster decision-making.	P close = 0.944 RMSEA = 0.042	Private

Yongda [3]	2024	UTAUT model, shortlisting models	The model predicts risks in digital HRM through risk event chains and graph.	N/A	N/A
Deviprasad et al. [4]	2023	ML algorithms	It uses risk event chains and the graph-based analysis to predict the potential threats in digital HRM.	FNR = 94.21 FPR= 89.62	Private
Daryani et al. [5]	2022	SVM, KNN, Word2Vec, Cosine similarity	Increases productivity per unit time and reduces amount of effort needed.	Accuracy = 91	Private
Satheesh et al. [6]	2020	Cosine similarity, NER, TF-IDF	The core advantages are efficiency increase, costs decrease, performance improvement, ability to foresee and forecast rising trends.	Cosine similarity = 0.6802	Private
Ulrich. [7]	1996	ML, Spacy NER	In the system, recruitment is simplified because it automates information extraction, makes screening more efficient, and ranks	N/A	Private

			candidates to ease and speed up hiring.		

CHAPTER 3

DESIGN METHODOLOGY

3.1 Introduction

The Automated HR System instantly integrates various stakeholders; applicants, administrators, employees, an email dispatch engine and an AI driven chatbot into a single perennial flow of recruitment and support stream. From the first second once a candidate uploads his résumé and basic profile information, the system's NLP powered parser converts unstructured text to actionable data that gets evaluated through a job criteria XGB-based ranking engine. The top candidates are automatically flagged and their records passed on to the administrator who has the CRUD control over vacancies, adjusts the scoring thresholds, and personalizes email templates. These customized messages irrespective of whether these are receipt acknowledgements, interview invitations or status updates are channels via a secure dispatch service which has a built-in delivery metrics model that can generate prompts as necessary. At the same time, the process is also modelled by its existing employees, who subscribe to a centralized notification hub for policy change or benefit update, and as an embedded chatbot, it works round the clock as an HR concierge: handling questions on everything from payroll dates to access to IT, with full account of the conversation and frequent referencing of its continuously updated knowledge base, upgrading complex cases to human specialists. By connecting all the facilities through common unifying data flows and intelligent automation, the system eliminates manual transactions and delays, makes uniform candidate and employee experiences, and lets HR experts shift their focus from ordinary administration towards high-level initiatives.

3.2 Project Architecture

The image shows a full end-to-end architecture of the AI-powered engine for HR automation, including processing, modeling and deployment of the resume data and chatbot interaction data in order to smooth recruitment and employee support activities. It starts with two types of data being presented by several companies (Company 1 to Company n). resume information from job seekers and chatbot data that consist of employee questions and HR communications. Both data types undergo standard pipelines of preprocessing, where permanence of data redundancy elimination, data cleaning eliminating inconsistencies, training and test sets are created, and textual property is tokenized into machine-readable formats is involved. After preprocessing, the data diverts into two different training modules. The first one is for the resume ranking model that is usually trained using the machine learning algorithms such as XGB, TF-IDF. This model learns to score and rank resumes according to the experience and skills and the needs of the job. The second module is training of an intelligent chatbot, perhaps on state-of-the art NLP models, who later on will answer employee HR-related questions. Both models are evaluated against learning criteria to acquire the adequate training, and as the criteria are met, the models are stored and validated by separate datasets to prove their efficiency.

When these models are trained and tested successfully, they are deployed. The resume ranking engine is connected to an online admin portal that enables HR managers to upload job descriptions, view the rank ordered candidates, and start email communication. The chatbot is delivered as a conversational UI and API to allow employee engagements for real-time query resolution. The system also has this unending monitoring of accuracy; if the performance falls below a certain limit, there is a step of reset to retraining with new data, thus, the model will adjust to the evolving trends and become better with each cycle. Such architecture is representative of an agile, intelligent and self-improving system that reduces human inputs when it comes to repetitive HR processes, improves accuracy of resume shortlisting, and provides effective employee assistance through automation. The visualization project architecture is shown in figure

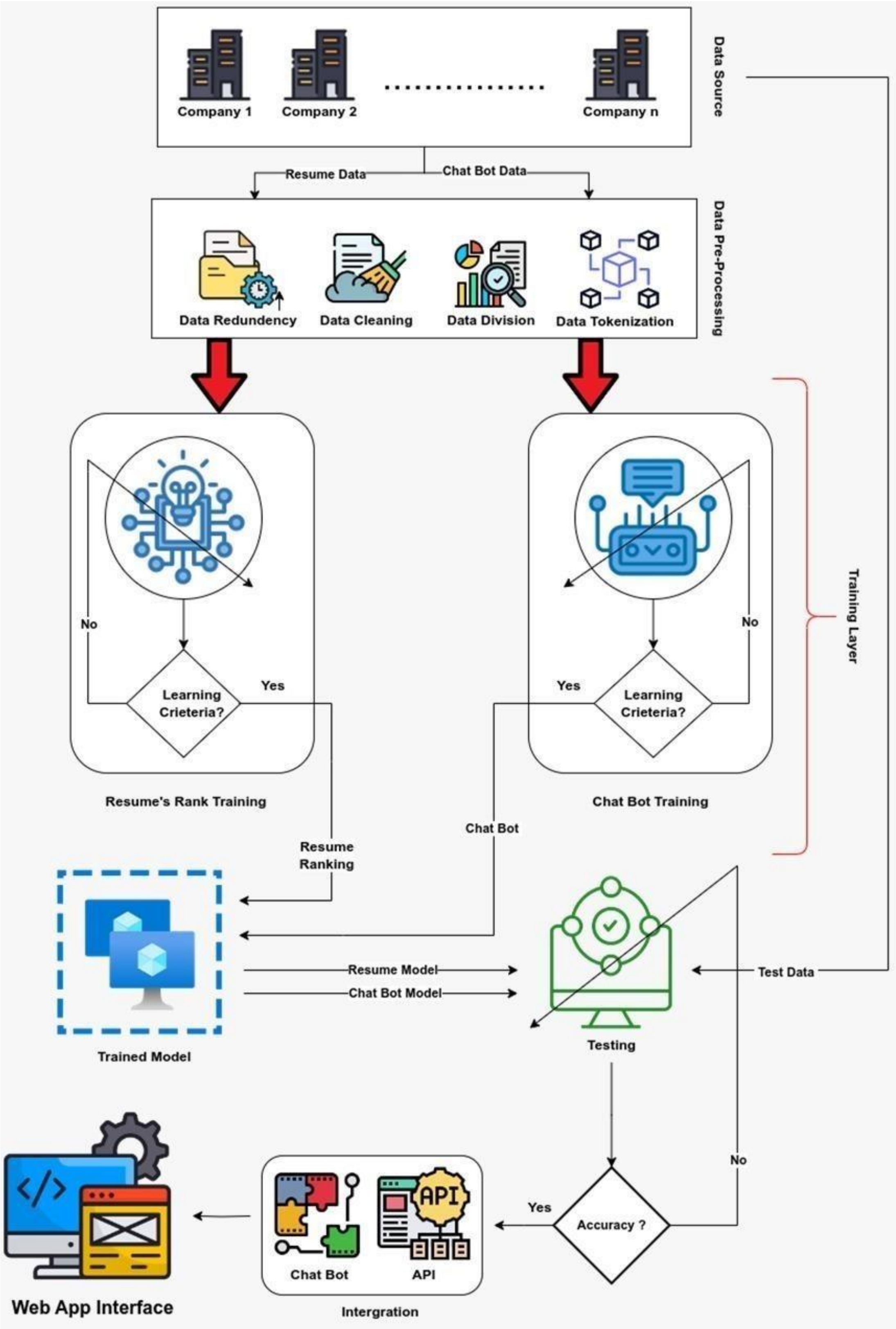

Figure 3. 1: Project Architecture

3.3 ERD Diagram:

The Human resources wing of any organization is required to play a double role which entails attraction of best talent through proper scrutiny of high-quality resumes and dealing with selected individuals to employ them or handling needs and concerns of current employees. A healthy HR management system gears companies to be at the top of their performance levels since the company does not only employ the right target, but also strives to create clear blue prints of the policies and procedures to success of the employee. On the contrary, the old, manually driven - HR processes are handicapped by the tedious processes that are time consuming, error prone and inefficient. Such period-dated workflows are also likely to slow down the decision making process, and things are likely to have adverse effects on the efficiency of HR function. Automating repeatable activities like resume grading, communication with the candidates, and implementation of policies, new systems give HR experts more breathing and thinking room for strategic moves and thus, making the organization more agile, more accurate, and people-focused. The whole process is shown in Figure
3. 2.

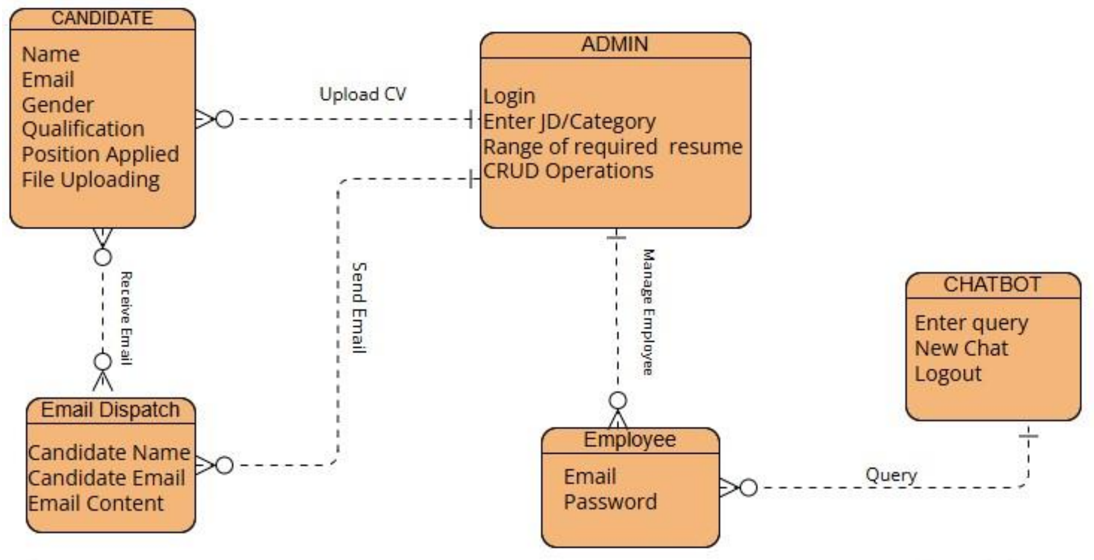

Figure 3. 2: Entity Relationship Diagram (ERD)

3.4 Project Flow

The flow of the "Automated HR" system is based on Home Page, which is the main access point with three main modules. the Candidate Page, the Admin Page and the Employee Page. Everyone of these modules is curated to certain types of roles play and functionality in the ambit of recruitment and HR workflow. Through the pages of candidates it will be possible to enter basic information of applicants, to upload their resumes and to receive e-mail based acknowledgements that they have applied. The Admin Page has two main functionalities, and these are the following: resume evaluation and management of employees using CRUD operations. Admins can view incoming CVs, send emails about tests or interviews, and work with the records of employees by executing Create, Read, Update, and Delete operations. Employee page allows the staff to log in securely with their credentials and unlock such options as submitting queries to the chatbot or viewing announcements or HR-related news as well as logging out of the system. Such an interface design that is modular and role based guarantees smooth, effective and secure flow of both information and interaction of candidates, administrators and employees, which improves communication and productivity throughout the entire HR process. The visualization of project flow is shown in figure 3.2.

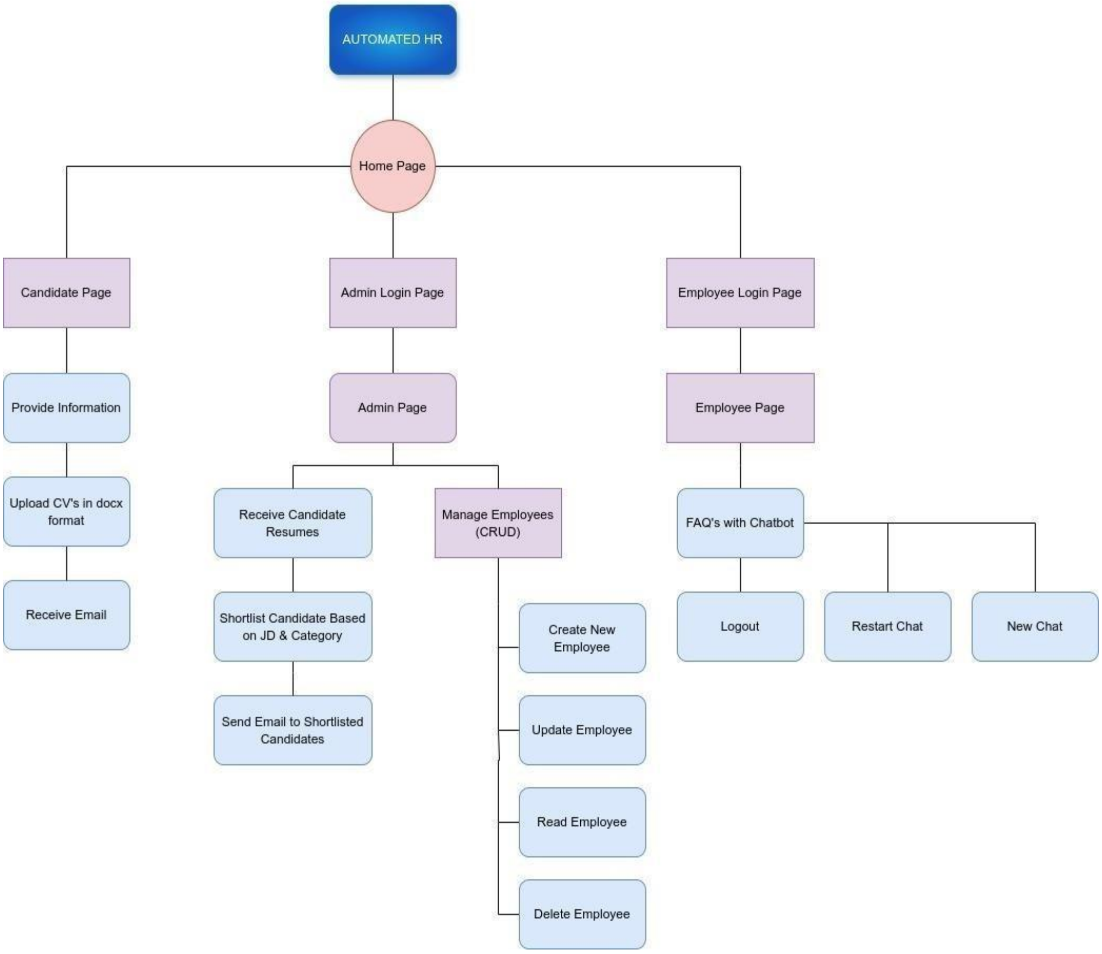

Figure 3. 3: Automated HR flow Diagram

3.5 Data Pre-processing

The first phase of the system involves gathering a vast number of resumes from different industries like IT, Finance, Quality Control in different companies. At the same time, the information for the chatbot training data is also collected through the most asked questions by employees' point of view. This wide and all-rounded dataset will ensure that the system can manage various job titles and employee questions effectively.
To be ready for the training of the models a swath of Natural Language Processing (NLP)-based preprocessing steps is performed:

- Elimination of Null Values: Incomplete or missing entries are eliminated for the sake of ensuring data reliability.
- Removal of Duplicate Entries: Unnecessary data is recognized and eliminated for the sake of consistency and bias avoidance.
- Tokenization: Text data is decomposed into small meaningful units (tokens) in an attempt to make the model understand and learn faster.

Using these preprocessing techniques, the system guarantees that only clean, structured and high-quality data is used to train AI models, in the aspect of accuracy and performance at later stages.

3.6 Resume Ranking Model Training

In the automated HR system, the use of the machine learning techniques is tactically used for an intelligent model of ranking a resume (CV). This model is oriented towards evaluating resumes as well as prioritizing them on several parameters including a candidate's skill set, professional experience, educational background, and the general fit of a person's resume for the job description. In the training phase, the system learns over a massive amount of historical resume-related data, extracting relevant features, such as keywords, roles, and context relevance, in the form of natural language processing (NLP). The training is iterative that is the model is constantly being perfected and tuned to increase the prediction and rank accuracy over time.
After training, the performance of the model is measured by means of standard classification metrics. Accuracy helps to inform the rate at which the model can correctly identify good resumes. Precision estimates the number of the resumes which are actually relevant to the job criteria, which helps to minimize false positives. Recall measures how much of the resumes that are really relevant to the position the model could find, so that good candidates would not be neglected. F1-score is a harmonic mean of precision and recall, a balanced take of how the model performs if precision and recall

are approximately equal. These assessments make sure that the model is not only giving good performance on training records but is also able to predict on the new unseen records accurately. Using this method, the system greatly increases the speed, accuracy, and equity of the recruiting process allowing the HR departments to use their efforts for more strategic hiring decisions, instead of manual screening.

3.7 Extreme Gradient Boosting model

XGB is a deep learning algorithm that builds a forest of decision trees for making predictions. It is able to classify tasks within a short time like predicting job types on the basis of a resume text. XGB increases the accuracy by adding decision trees sequentially, so that each tree corrects the mistakes of the previous ones. It is both highly performant and fast and supports regularization, which makes it especially good for dealing with big or high-dimensional data, such as TF-IDF vectors from resume files.

XGB uses objective function having loss function and an extra regularization term as shown below:

$$L(\varphi) = \Sigma\ [l\ (y_i, \hat{y}_i^{(t)})] + \Sigma\ [\Omega(f_k)]$$

Where:

- $L(\varphi)$ is the sum of the objective function.
- $l\ (y_i, \hat{y}_i^{(t)})$ is the loss function.
- $\Omega(f_k)$ is the regularization term.
- $f_k(x) = w\_q(x)$ is the k-th decision tree prediction score.

The regularization term prevents the problem of overfitting and may be represented as:

$$\Omega(f) = \gamma * T + (1/2) * \lambda * \Sigma\ (w_j^2)$$

Where:

- T is the leaf count of trees.
- w_j is the j-th leaf weight.
- γ is the cost of additional leaf node.
- λ is the L2 regularization coefficient.

XGB maximizes the objective function in the second-order Taylor expansion (including Hessian and gradient) that accelerates the convergence and eases the splits. In multi-class classification for instance, in order to bring back the classification into categories (like Data Science or HR), the SoftMax function is used to estimate class probabilities.

$$P(y_i = k) = e^{\wedge}(\hat{y}_{ik}) / \Sigma\, e^{\wedge}(\hat{y}_{ij}) \text{ for } j = 1 \text{ to } K$$

Inputs for XGB are resume texts, which are initially converted to a numerical value using methods like TF-IDF (Term Frequency-Inverse Document Frequency). Numerical vectors provide an objective indication of the importance of each word of a resume and make it possible for the model to find words pertinent to some job profile.

Statistics that are used to estimate the model performance are:

- **Accuracy**: (Correct Predictions / Total Predictions)
- **Precision**: TP / (TP + FP)
- **Recall**: TP / (TP + FN)
- **F1-score**: 2 * (Precision * Recall) / (Precision + Recall)

3.8 Chatbot Training

Apart from the resume rank model, the system also learns a chatbot model that is able to respond to the new hire questions. Chatbot model is induced on the basis of FAQs and generic HR-related questions. It has the same learning cut-off with the resume model so it will provide response that is correct and contextual.

The chatbot is routinely trained and tested with a view of sharpening its conversational skills hence it would be a useful tool for employee on-boarding.

3.9 Accuracy Evaluation

After training, the chatbot and CV ranking both models are tested on unseen data. The system measures their performance using the same accuracy metrics as given before. If the models fail to pass the tests of accuracy, they are resubmitted to undergo retraining. Otherwise, they are carried into API integration stage after having extensive optimizations in advance, before deployment.

3.10 Web Application Integration

The trained models are effortlessly incorporated in the HR web application through an API architecture. This API-system guarantees the effective exchange of information between the backend machine learning models and the frontend web interface facilitating flow of data and providing real-time interaction.

- The resume ranking model is an essential component of the HR portal because it is the one that automates the shortlisting process for candidates. After uploading of resumes, the model then places them in order according to various criteria ranging from skills, experience, and education. Through this automation, the review of resumes is made at much lower timeframes so that only those resumes which are most relevant to the job are taken into consideration for further assessment. Applying this model, an HR team can focus on such higher-level functions as arranging an interview schedule and communicating with the candidates.

- The chatbot model is embedded in the system for real time interactivity in engaging the employees. It may help to provide the answers to the frequently asked questions, lead employees through the application process or give the information on HR policies and procedures. This 24/7 support guarantees that the employees are in touch with information and help right away which enhances their experience and satisfaction with the organization.

By including these models, the HR system becomes effective; it is efficient in that it eliminates the administrative overhead, there is improved accuracy, and decisionmaking is enhanced. The integration of AI-powered tools in the workflow of HR is not only the feature to support user experience, but it also simplifies a number of HR operations, from candidate selection to the support of employees. This will make the HR process more slimline and effective, enabling the HR members of the staff to concentrate on strategic tasks that would bring more value to the organization.

CHAPTER 4

DATA AND EXPERIMENTS

4.1 Data Description

The information in this project has 5,744 rows and 2 primary columns, which primarily include candidate CVs to sort and classify. The two primary columns are "Category" and "Resume." The "Category" column represents the occupation category or profession of a resume, like data science, HR, or software engineering. With 4,717 non-null values, it is evident that there are some rows with no occupation category mentioned. The "RIsum" column keeps the whole text data of the resumes, including education, skills, interests, and project experiences. It contains 4,757 non-null values, i.e., there are virtually no missing resumes in the data set.

This information is extremely helpful for many Natural Language Processing (NLP) applications in Human Resource Management such as resume categorization, skill extraction, candidate-job matching, and resume screening. Since every resume is labelled as a category, the information can be used in supervised learning models to predict job categories based on text information.

However, the unstructured form of the resume data is difficult and challenging. Preprocessing operations such as cleaning, tokenization, and feature extraction are to be done on the data to prepare it for processing. Feature extraction can be done through operations such as TF-IDF, Word2Vec, and Stop Word removal. Even the presence of missing values in the data requires null entries to be processed and resume formats to be normalized before machine learning models can be applied.

All in all, this data provides a good real-world environment in which to test AI-based HR gear and studies on AI-aided employment gear, such as filtering, sorting, and ranking of resumes. It is a good place to develop automated tools for automating HR and optimizing the efficiency of hiring processes.

4.2 Feature Extraction Techniques

Feature extraction is a critical step of converting free-form resume text into structured information readable by machine learning algorithms. Within the data, the "Resume" column contains raw, free-text accounts of a candidate's education, work experience, skills, and other professional credentials. Since this is information in natural language, it needs to be translated into numeric or categorical format using a series of Natural Language Processing (NLP) algorithms.

To machine learn resume data for use in classification and recommendation in Human Resource Management (HRM) systems, several of the most important feature extraction techniques are utilized. The first step towards this is text preprocessing, where the text from the resume is broken down into smaller, more informative units of data. This helps in pre-processing the text data into a format that is favorable for additional analysis so that algorithms can process and understand the content adequately. The sample cleaned data is shown in figure 4.1.

```
                                       Cleaned_Resume
0     areas of interest deep learning control system...
1     skills r python sap hana tableau sap hana sql ...
2     education details mca ymcaust faridabad haryan...
3     skills c basics iot python matlab data science...
4     skills python tableau data visualization r stu...
```

Figure 4. 1: Sample of cleaned resume data

Tokenization divides the text into words or tokens, and removal of stop word removes common but non-informing words like “and”, “the” and “in” that do not facilitate classification. Lowercase makes it normal to become lower case throughout the text. Other data normalization, stemming or lemmatization brings a word to its root form i.e. bring “running” to “run” in a way that words with different forms of the same word are treated as the same as a single feature as shown in figure 4.2.

```
                                               tokens
[area, interest, deep, learn, control, system,...
[skill, r, python, sap, hana, tableau, sap, ha...
[educ, detail, mca, ymcaust, faridabad, haryan...
[skill, c, basic, iot, python, matlab, data, s...
[skill, python, tableau, data, visual, r, stud...
```

Figure 4. 2: Sample of tokens of resume text

After pre-processing of the text, logic tells us to then be numerically converted through vectorization method. TF-IDF (Term Frequency-Inverse Document Frequency) is one amongst which the words are arranged frequency and even pertinence wise. The word frequencies are de-ranked and place more emphasis on the individual resume or job title words of a higher relevance.

Besides the standard text features, the specific ones may be drawn from the ordered or semi-order parts of the resume. Such are quantitative characteristics, e.g. the number of years of working as a professional, the level of education (bachelor’s degree, master's degree), volume of technical knowledge, relative frequency of presence of some of the keywords like "Python," "Excel," or "project management.". These specialized attributes can be numerically represented and appended with TF-IDF vectors to create an overall set of features for training a model.

Finally, Category column is the variable that is to be labelled and that contains the following tags as its elements: Data Science, HR or Software Engineering. These tags have to be converted in numeric form with the help of Label Encoding or One-Hot Encoding based on the Machine learning model. Label encoding maps categories to an integer but, when it comes to encoding categorical values, one-hot encoding provides binary vectors.

Through these feature extraction steps, the resume database is converted to a repository which can be used in applying the supervised learning algorithms. This makes it possible to automate such tasks as resume classification, job matching, and smart filtering – fundamental components of the AI-based HRM systems of the contemporary world.

CHAPTER 5

RESULTS AND DISCUSSIONS (or USER MANUAL)

5.1 Results of different model

The comparison between four classification models—Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGB suggests that XGB is the best among all four models according to all the metrics of evaluation. While Logistic Regression was good with an accuracy score of 0.764 and an F1 score of 0.790, the others outperformed it. Random Forest does much better than this, with accuracy 0.862 and F1 of 0.880, showing better precision and recall balance. SVM does similarly, with slightly lower accuracy (0.847) but with a comparable F1 score of 0.877, showing its robustness against imbalanced classes. However, XGB shows the highest accuracy (0.888), precision (0.925), recall (0.901), and F1 score (0.905) out of these models and thus is the most efficient model in this comparison. The result indicates the conclusion that XGB is the strongest and most stable model appropriate for classification work here as shown in table 5. 1.

Table 5. 1: Results of classifiers models

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	0.764	0.832	0.784	0.790
Random Forest	0.862	0.920	0.876	0.880
Support Vector Machine	0.846	0.910	0.866	0.877
XGB	0.88	0.924	0.901	0.904

We can clearly visualize the improvement before and after implementing the techniques in figure 5.1.

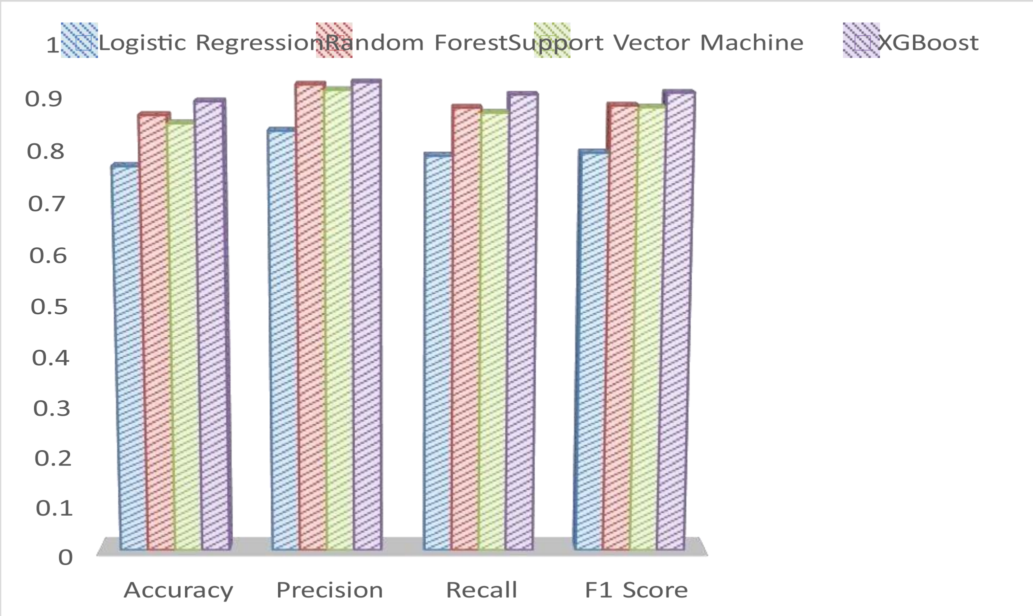

Figure 5. 1: Chart of classifiers models

Performance of regression model comparison indicates that the ensemble-based models, i.e., Random Forest Regressor and XGB Regressor, exhibit improved performance as opposed to the standard and linear models in the prediction of values. Linear Regression is the poorest performing with worst Mean Squared Error (MSE) being 136.12, Root Mean Squared Error (RMSE) of 11.67, and worst R^2 score of 0.19, which reflects poor model explanatory power. The Support Vector Regressor does a bit better but is still lacking, with an MSE of 119.84, RMSE of 10.95, and R^2 score of 0.29. Compared to the Random Forest Regressor does a much lower MSE of 44.43, RMSE of 6.67, and a very high R^2 score of 0.736, indicating that it does a much better job of modelling the variance in the data. XGB Regressor is also very close with an MSE of 44.61, RMSE of 6.68, and an R^2 score of 0.735, but better than Random Forest on Mean Absolute Error (MAE) with the best MAE of 3.81. These indicate that both the ensemble models are very good, but XGB has an advantage of being able to predict better results according to table 5. 2.

Models	**MSE**	**RMSE**	**MAE**	**R2 Score**
Linear Regression	136.11	11.66	8.11	0.19
Random Forest regressor	44.42	6.66	4.24	0.73
Support Vector Regressor	119.83	10.94	8.63	0.28
XGB Regressor	44.61	6.67	3.80	0.73

Table 5. 2: Results of regressor models

We can clearly visualize the improvement before and after implementing the techniques in figure 5.2.

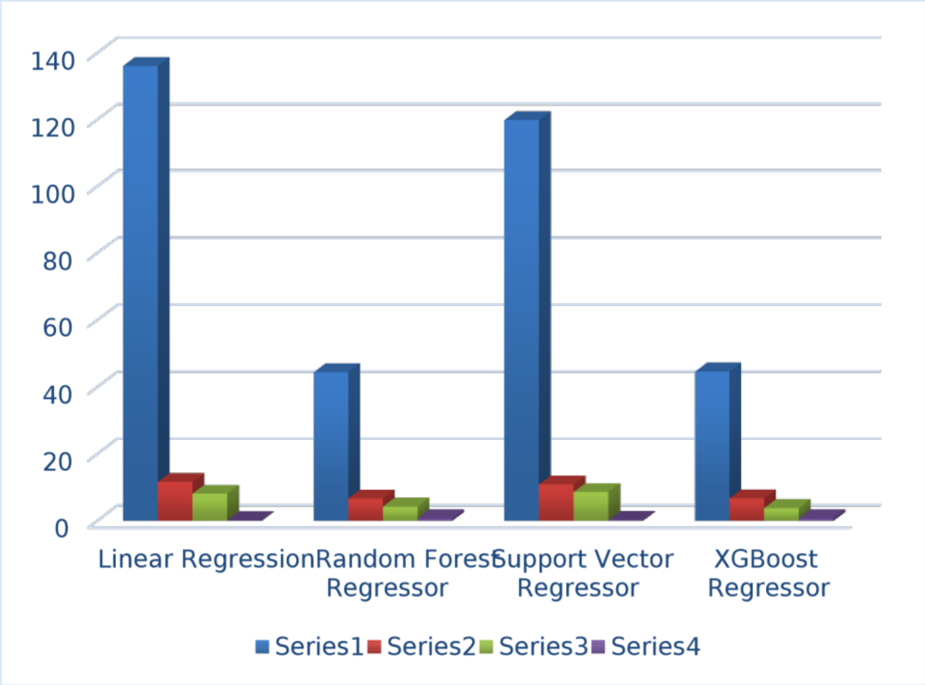

Figure 5. 2: Chart of regressor models

5.2 Bar chart

The aggregate resumes received under different categories of jobs that leads to imbalance of data distribution. The maximum number of resumes are lead by sales, HR and information technology and all of them exceed 230 which implies higher availability or interest of candidates in these sections. On the other hand, such posts as Civil Engineer, SAP Developer and Network Security Engineer have fewer resumes with most Job seekers being less than 50, reflecting underrepresentation. This bias means that any machine learning algorithm that is trained on this data set will also tend to become biased towards majority classes and find it hard with minority ones, leading to overfitting and poor generalization. This bias must be addressed so that a proper and fair model can be trained. The distribution of category over resume is shown in figure 5.3.

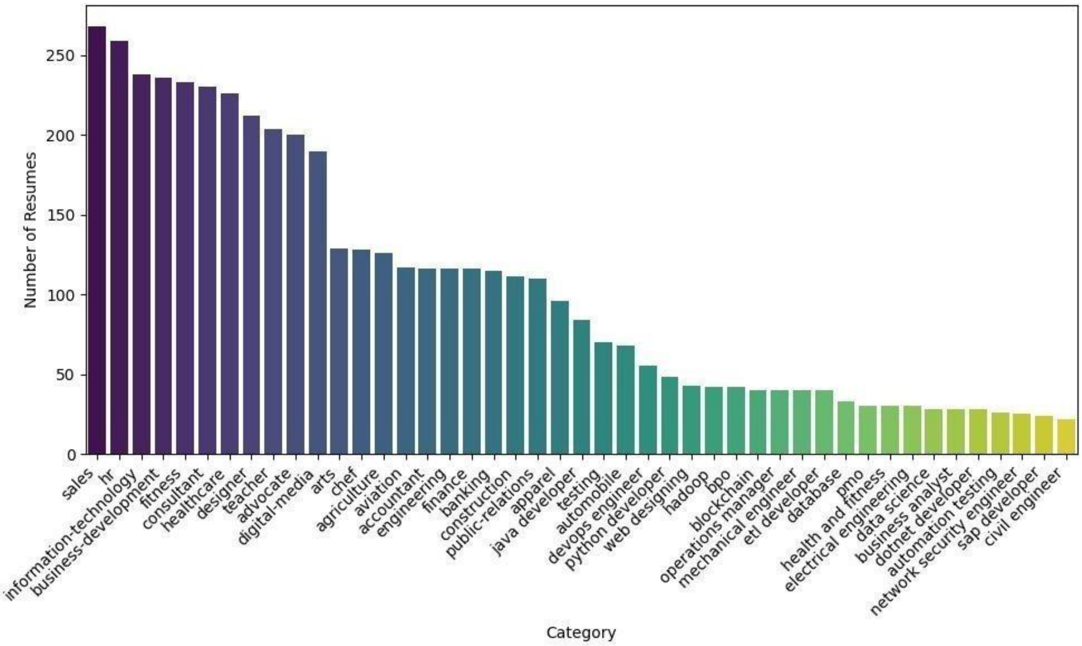

Figure 5. 3: Distribution of resumes by category

5.2.1 Confusion matrixes

Two confusion matrices show the performance of the classifier on training and the set of tests for 45 classes (labelled 0-44). In the training matrix, almost all cells line up along the main diagonal – many of them having values of 80 and even higher, up to the 190 – meaning that nearly all the training samples were correctly classified to their true classes. This emphasized diagonal dominance reveals amazingly high accuracy of the training and points to the fact that the model has memorized the training data well. On the other hand, while the test confusion matrix still has a diagonal dominance, its diagonal components are mostly lower (below 70) and has more off-diagonal entries. Misclassifications, made by this model when facing the unseen data, demonstrate that the model commits more mistakes and sometimes mixes up the pairs of classes. However, the persisting diagonal dominance shows that the model continues having a good chance of predicting new samples with reduced precision from those on the training set. The confusion matrixes of both testing and training are in figures 5.4 and 5.5.

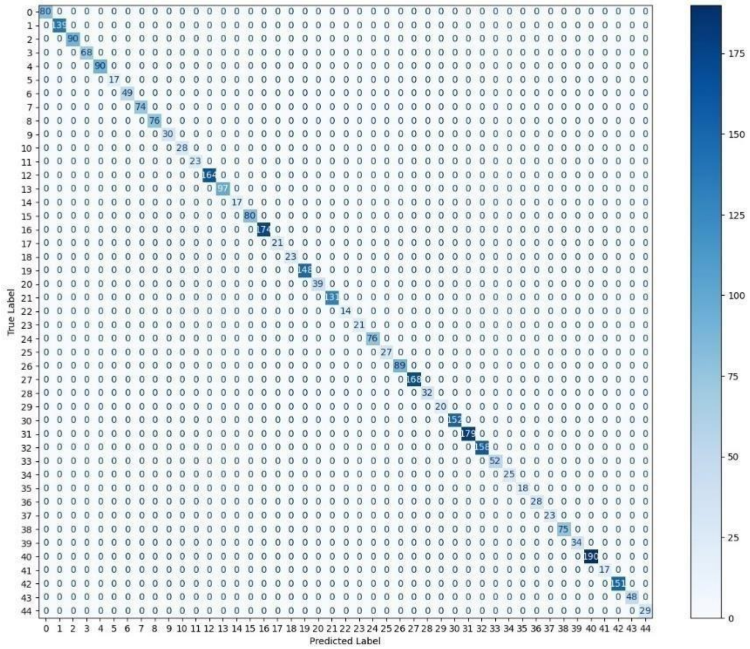

Figure 5. 4: Training confusion matrix

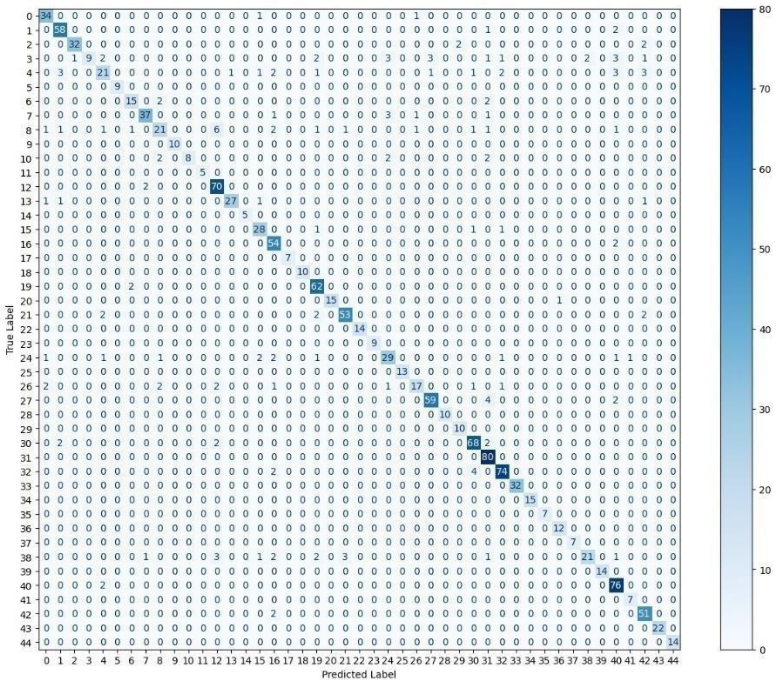

Figure 5. 5: Testing confusion matrix

5.3 Home page

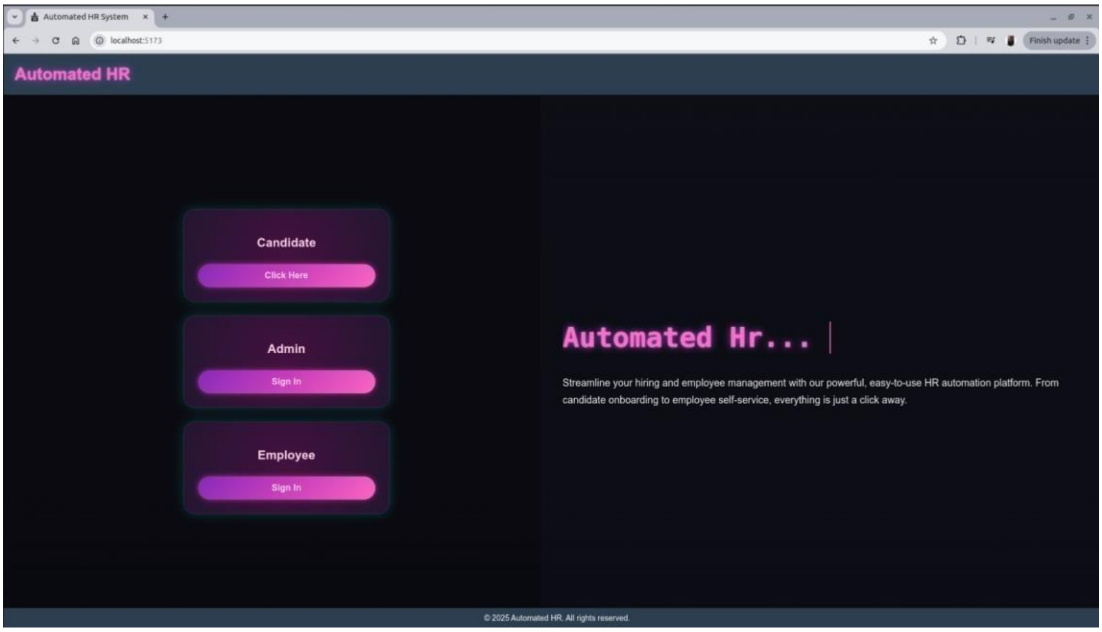

Figure 5. 6: Home page

5.4 Candidate page

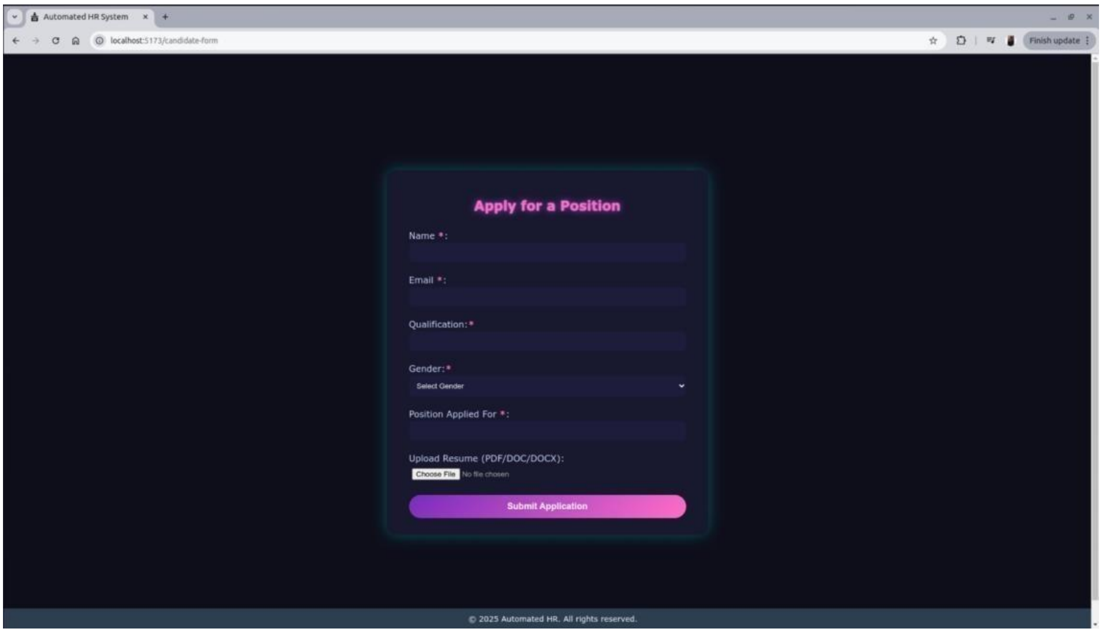

Figure 5. 7: Candidate page

5.5 Admin login page

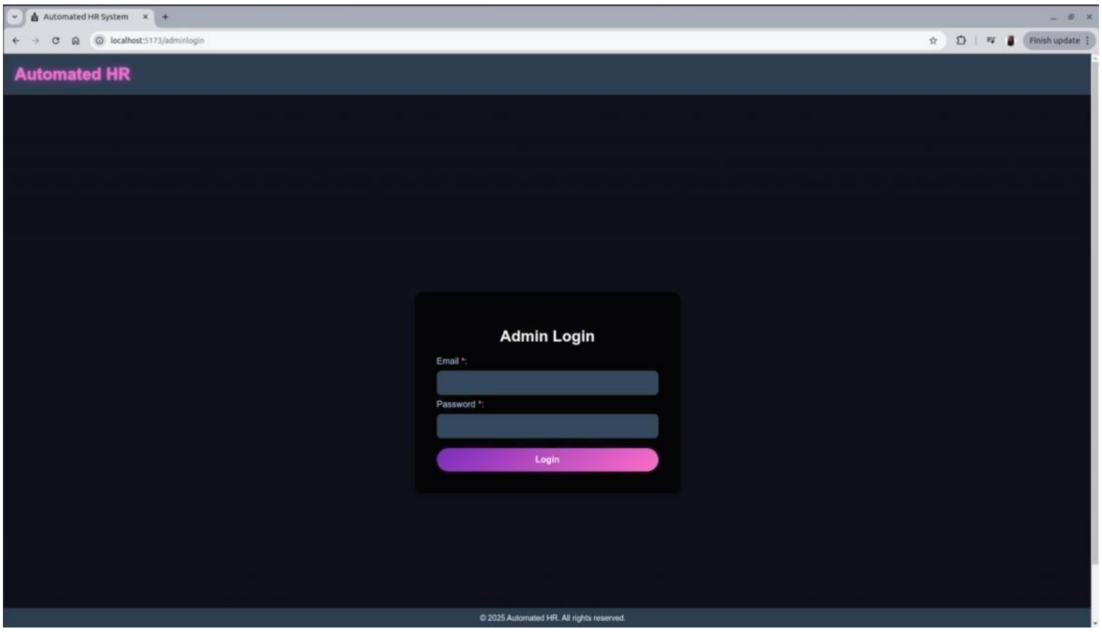

Figure 5. 8: Admin login page

5.5.1 Admin panel

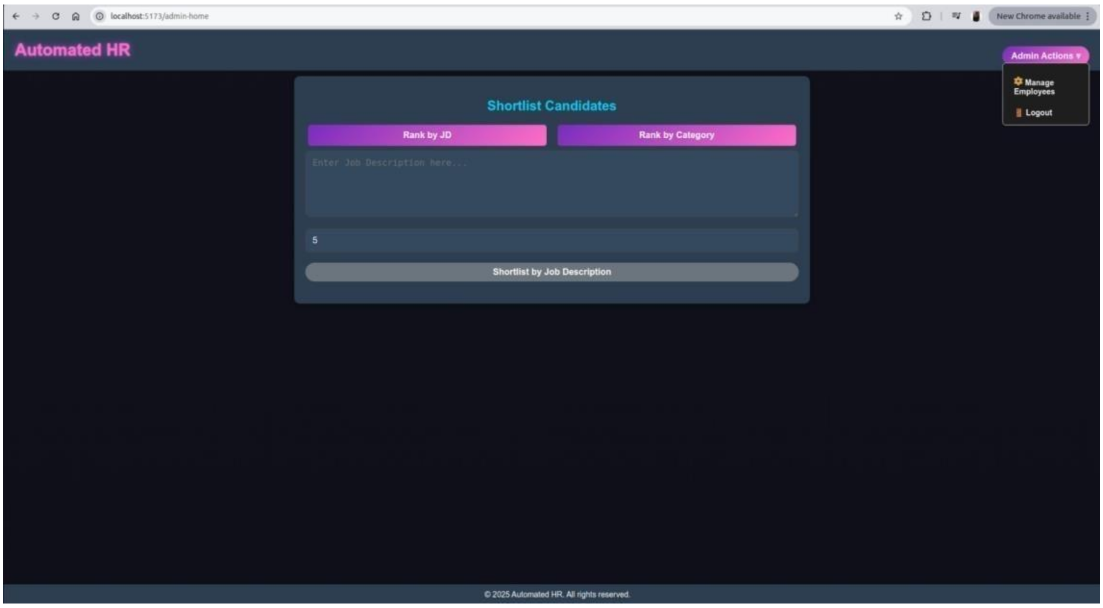

Figure 5. 9: Admin panel

Figure 5. 10: Email receiving to selected candidate

5.6 Employee

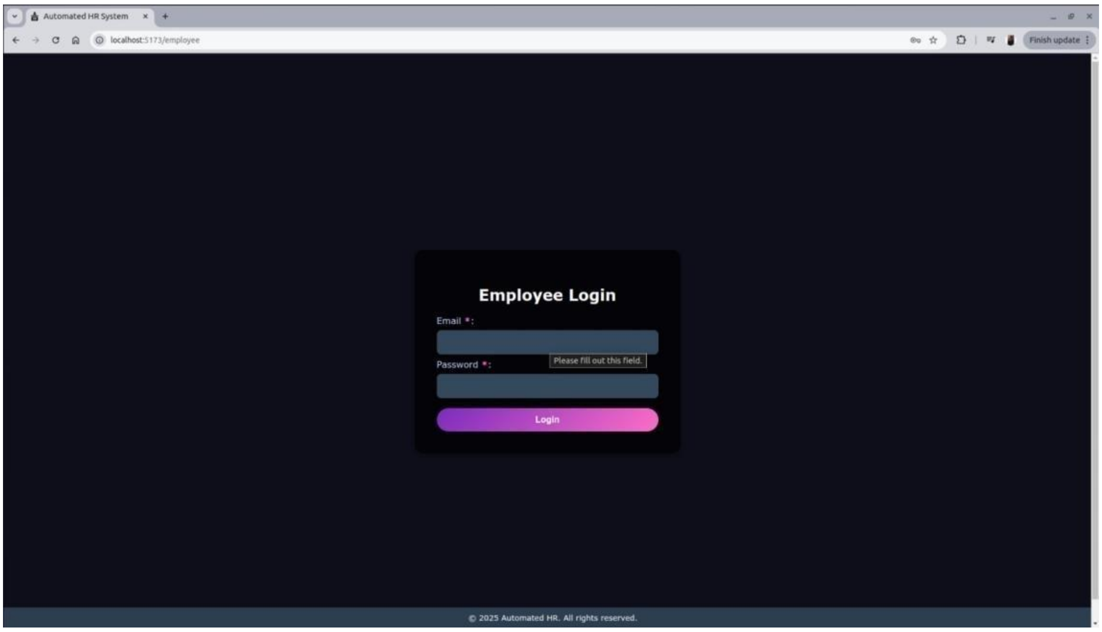

Figure 5. 11: Employee Login

5.7 Chatbot

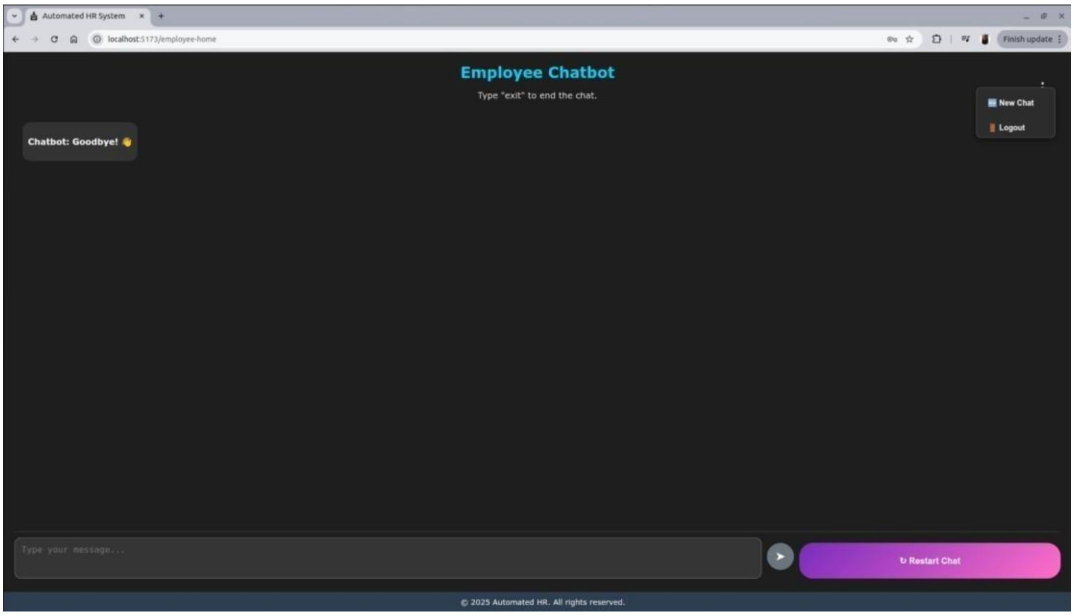

Figure 5. 12: Chatbot interface

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

The AI-powered recruitment platform uses cutting technologies, such as Natural Language Processing, Machine Learning models, for example, XGB, and textanalysis algorithms like TF-IDF to automate HR standard activities and increase efficiency by several times. With more than 5K real-world resumes mined from different industries (IT, finance, HR, quality control, and so on), the system gets trained to scan titles, estimate experience levels, and pick out required skills to generate a sorted shortlist, which would spare people hundreds of hours. Lean development pipeline is six-step: data collection, model training, performance evaluation, web integration, scalability, and accuracy. Metrics during evaluation such as precision, recall, and the general accuracy validate that only the best qualified candidates make it through to becoming software engineer or data analyst. Besides screening, the platform has a real-time HR chatbot for new hires and permanent staff walk through the onboarding process and simple questions without involving the HR staff, which saves the latter from repetitive questions. Automated email workflows including provisions for reminders for interview scheduling and follow-ups are effortlessly choreographed through a web portal and RESTful API, enabling hiring managers to have one dashboard to post openings, view ranked applications, and communicate with candidates. With end-to-end, data-driven recruitment processes where each step of recruitment – from resume ranking to onboarding – is converted into an end-to-end, data-driven process, the system cuts costs and turnaround times, enabling HR teams to employ a strategic approach instead of administrative overhead.

6.2 Recommendations:

In order to improve the performance in addition to the scalability of the AI-driven recruitment system even further, several improvements are suggested. By broadening the training corpus to cover a wider range of industries and positions, the model will be able to generalize better to different kinds of hiring settings and lessen bias resulting from positions that are overly represented. When doing the data prep, using reliable methods for imputing missing fields and approach normalization and outlier management with care will guarantee that predictions are accurate in underrepresented job types. Scaling up of the conversational engine to make use of transformer-based architecture like BERT or GPT will equip the chatbot with enhanced, contextual understanding, giving more natural and correct answers at every step of onboarding and employee support. Imprinting a closed-loop feedback mechanism in which applicants and HR staff can evaluate the relevance of candidate matches or chatbot exchanges will ensure a constant flow of real-world inputs for iteratively improving the ranking and dialogue models. In order to protect sensitive information, the use of role-based access control (RBAC) will define administrators, HR, and general staff permissions to ensure that each user only has access to what they are supposed to perform and of the data available. As for the analytics, dynamic dashboard integration will enable HR leaders to track system health in a real-time fashion – meaning monitoring metrics like resume throughput, delivery rates on email, statistics on chatbot resolution – while continuous retraining of models off of newly ingested resumes and feedback will align the system with the changing state of the labor market. Lastly, the use of specialized resume-parsing APIs can enhance the immense extraction of structured attributes –skills, qualification and employment history from a wide variety of CV formats thereby honing the screening accuracy of the system. Together, these improvements will take the platform and turn it into a genuine enterprise-class solution – secure, interactive to the core, and continuously learning of its own usage.

REFERENCES

[1] P. Yang and J. Li, "Construction and application of intelligent human resource management system based on machine learning algorithm," *Journal of Computational Methods in Sciences and Engineering,* p. 14727978251318814, 2024.

[2] N. Nawaz, H. P. B. K. Arunachalam and V. Gajenderan, "The adoption of artificial intelligence in human resources management practices," *International Journal of Information Management Data Insights,* vol. 4, p. 100208, 2024.

[3] Hu and Yongda, "Risk Prediction of Digital Human Resource Management Based on Artificial Intelligence," *Journal of computing and information technology,* vol. 30, pp. 23--33, 2022.

[4] Deviprasad, S. a. Madhumithaa, N. a. Vikas, I. W. a. Yadav, A. a. Manoharan and Geetha, "The machine learning-based task automation framework for human resource management in MNC companies," *Engineering Proceedings,* vol. 59, p. 63, 2023.

[5] C. a. C. Daryani, G. S. a. Patel, H. a. Chhabra, I. K. a. Patel and Ruchi, "An automated resume screening system using natural language processing and similarity," *ETHICS AND INFORMATION TECHNOLOGY [Internet]. VOLKSON PRESS,* pp. 99--103, 2020.

[6] K. a. J. Satheesh, A. a. Iswarya, L. a. Ayesha, K. a. Bhanusekhar, G. a. Hanisha and K, "Resume ranking based on job description using SpaCy NER model," *International Research Journal of Engineering and Technology,* vol. 7, pp. 74-77, 2020.

[7] Walwei and Ulrich, "Improving job-matching through placement services," *International handbook of labour market policy and evaluation,* 1996.

[8] "An extended UTAUT model for analyzing users' Acceptance factors for artificial Intelligence adoption in human resource recruitment: A case study of Thailand".

- -

-

Document Details

Submission ID trn:oid:::1:3248784640

Submission Date

May 13, 2025, 5:04 PM GMT+5

Download Date

May 13, 2025, 5:15 PM GMT+5

File Name

FYP_Final_Report_without.pdf

File Size

1.3 MB

34 Pages

6,709 Words

36,797 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Frequently Asked Questions

How should I interpret Turnitin's AI writing percentage and false positives?
The percentage shown in the AI writing report is the amount of qualifying text within the submission that Turnitin's AI writing detection model determines was either likely AI-generated text from a large-language model or likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

False positives (incorrectly flagging human-written text as AI-generated) are a possibility in AI models.

AI detection scores under 20%, which we do not surface in new reports, have a higher likelihood of false positives. To reduce the likelihood of misinterpretation, no score or highlights are attributed and are indicated with an asterisk in the report (*%).

The AI writing percentage should not be the sole basis to determine whether misconduct has occurred. The reviewer/instructor should use the percentage as a means to start a formative conversation with their student and/or use it to examine the submitted assignment in accordance with their school's policies.

What does 'qualifying text' mean?
Our model only processes qualifying text in the form of long-form writing. Long-form writing means individual sentences contained in paragraphs that make up a longer piece of written work, such as an essay, a dissertation, or an article, etc. Qualifying text that has been determined to be likely AI-generated will be highlighted in cyan in the submission, and likely AI-generated and then likely AI-paraphrased will be highlighted purple.

Non-qualifying text, such as bullet points, annotated bibliographies, etc., will not be processed and can create disparity between the submission highlights and the percentage shown.

turnitin Page 2 of 36 - AI Writing Overview Submission ID trn:oid:::1:32487846

APPENDICES

ORIGINALITY REPORT

12%	2%	1%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

Submitted to Skyline University College

1	Student Paper	10%
2	www.jetir.org Internet Source	<1%

3	assets.cureusjournals.com Internet Source	<1 %
4	www.mdpi.com Internet Source	<1 %
5	Submitted to UCL Student Paper	<1 %
6	discovery.researcher.life Internet Source	<1 %
7	arxiv.org Internet Source	<1 %
8	ijsrset.com Internet Source	<1 %
9	"Classification in BioApps", Springer Science and Business Media LLC, 2018 Publication	<1 %
10	ia801504.us.archive.org Internet Source	<1 %
11	pt-corp.com Internet Source	<1 %
12	www.customerservicemanager.com Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

-

GRADEMARK REPORT

GENERAL COMMENTS

FINAL GRADE /0

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22

PAGE 23

PAGE 24

PAGE 25

PAGE 26

PAGE 27

PAGE 28

PAGE 29

PAGE 30

PAGE 31

PAGE 32

PAGE 33

PAGE 34