DATA SCIENCE-BASED STAFF RECRUITMENT ANALYSIS AND PREDICTION

Dr.s.sridhar [1], Gangu Naveen [2]

[1]Associate Professor,Department of MCA, Miracle Educational Society Group Of InstitutionsBhogapuram Mandal, Vizianagaram, Andhra Pradesh 535216

[2]PG Scholar , Department of MCA, Miracle Educational Society Group Of InstitutionsBhogapuram Mandal, Vizianagaram, Andhra Pradesh 535216

Abstract: This project was created for ACN Infotech (India) Pvt Limited using data-driven methodology. This project makes hiring staff or employees quicker and easier than it would be using the traditional hiring process. use machine learning and data analysis techniques to forecast and evaluate the potential outcomes of hiring new employees for a company. Using a variety of datasets, we examine the factors that influence hiring decisions based on specific characteristics such as job experience, skills, specialisation, academic standing, and percentage. To develop a predictive model that forecasts the potential outcomes of the hiring process, we employ a variety of machine learning algorithms, including feature selection, data cleaning, data preprocessing, model training, and model assessment. In order to have a comprehensive grasp of the dependent elements that the recruiting process depends on, we offer insights on staff recruitment. The information supplied aids the hiring team of companies or sectors in determining the potential outcomes of a candidate and moving forward with the hiring process. Based on a review of the applicants who applied for the position and those who were hired, this project also offers trends and patterns. The results give practical implications for improving the placement tactics as well as insightful information on the dynamics of staff recruiting. Furthermore, our solution offers thorough assessments of recruiting trends and patterns seen among applications and recruited individuals, going beyond simple prediction. These results provide insight into the dynamics of hiring new employees, pointing out patterns, pinpointing areas in need of development, and guiding future strategic placement plans. Our project's findings enable organisations' and industries' recruiting teams to make well-informed decisions at every level of the hiring process by offering them essential recommendations. Our research helps recruiters to efficiently prioritise prospects, allocate resources, and speed decision-making by identifying the elements that have a substantial impact on recruiting results. This, in turn, leads to more successful and efficient recruitment procedures. With the use of a variety of machine learning algorithms, we have created a predictive model that can accurately anticipate the outcomes of hiring. To ensure its resilience and efficacy in practical situations, this model is put through a rigorous procedure that includes data cleaning, preprocessing, feature selection, model training, and assessment.

Index terms - Staff Recruitment, Data Science, Machine Learning, Predictive Model, Data Analysis, Feature Selection, Model Training, Model Evaluation, Recruitment Trends, Candidate Prediction, Recruitment Strategy, Data Preprocessing, Hiring Process Optimization, Recruitment Insights, ACN Infotech.

1. INTRODUCTION

This project was created for ACN Infotech (India) Pvt Limited using data-driven methodology. This project makes hiring staff or employees quicker and easier than it would be using the traditional hiring process. use machine learning and data analysis techniques to forecast and evaluate the potential outcomes of hiring new employees for a company. Using a variety of datasets, we examine the factors that influence hiring decisions based on specific characteristics such as job experience, skills, specialisation, academic standing, and percentage. To develop a predictive model that forecasts the potential outcomes of the hiring process, we employ a variety of machine learning algorithms, including feature selection, data cleaning, data preprocessing, model training, and model assessment. In order to have a comprehensive grasp of the dependent elements that the recruiting process depends on, we offer insights on staff recruitment. The information supplied aids the hiring team of companies or sectors in determining the potential outcomes of a candidate and moving forward with the hiring process.

Based on a review of the applicants who applied for the position and those who were hired, this project also offers trends and patterns. The results give practical implications for improving the placement tactics as well as insightful information on the dynamics of staff recruiting. Furthermore, our solution offers thorough assessments of recruiting trends and patterns seen among applications and recruited individuals, going beyond simple prediction. These results provide insight into the dynamics of hiring new employees, pointing out patterns, pinpointing areas in need of development, and guiding future strategic placement plans. Our project's findings enable organisations' and industries' recruiting teams to make well-informed decisions at every level of the hiring process by offering them essential recommendations. Our research helps recruiters to efficiently prioritise prospects, allocate resources, and speed decision-making by identifying the elements that have a substantial impact on recruiting results. This, in turn, leads to more successful and efficient recruitment procedures. With the use of a variety of machine learning algorithms, we have created a predictive model that can accurately anticipate the outcomes of hiring. To ensure its resilience and efficacy in practical situations, this model is put through a rigorous procedure that includes data cleaning, preprocessing, feature selection, model training, and assessment.

2. LITERATURE SURVEY

2.1 "Indian IT services companies heighten candidate scrutiny, take longer to close the mandate. While demand is rebounding, particularly for specialized fields like AI and cyber security, the hiring process has become more rigorous, leading to prolonged recruitment timelines.", Published by Richa Bhattacharya, Times of India

ABSTRACT:

Rich Bhattacharya, a well-known data scientist, wrote this study in Time of India. The passage draws attention to a change in Indian IT services businesses' employment procedures, showing a tendency towards more stringent screening procedures and extended hiring periods. The resurgence in demand for specialised skills like cyber security and artificial intelligence (AI) is one of the reasons for this shift in strategy. Companies are taking a more stringent approach to applicant evaluation as a result of the renewed need for specialised domains like artificial intelligence and cyber security. As these specialised fields in the IT sector become more well-known, employers are giving more weight to hiring applicants who possess particular knowledge and skills in these fields. Companies have been forced to improve their hiring practices due to the heightened competition for talent in these specialised industries. Businesses are spending more time and money evaluating applicants' fit for the position as they compete for top personnel with in-demand abilities. In order to fully evaluate individuals' talents and organisational fit, this increased scrutiny may entail several rounds of interviews, technical tests, and evaluations. These reasons have caused the hiring process to take longer, which has resulted in longer closure mandate timelines. In order to make well-informed recruiting decisions and enrol applicants who not only have the necessary abilities but also fit with the company's culture and long-term objectives, businesses are taking extra care. Even though there is still a high demand for talent, businesses are putting quality above speed in their recruiting procedures because they understand how crucial it is to have the right people in place to spur innovation and competitiveness in a market that is becoming more and more dynamic.

2.2 "Startup recruiters find few takers amid overall hiring slowdown. In a harsh funding environment, where hiring has plummeted as startups focus on profitability, many recruiters have been let go as part of a cost-cutting exercise or are looking out for an option.", Published by Sreeradha Basu, Economictimes.

ABSTRACT: Vikram Bhagath published this article in the Times of India. The statement draws attention to a recent trend that has been seen by several businesses: a noticeable rise in the quantity of job applications received. Because they must filter through and assess a sizable pool of applicants in order to find qualified hiring, recruiters face a huge problem as a result of this spike in applications. The number of

people looking for jobs at different firms has suddenly and significantly increased. There are a number of reasons for this rise in the number of applications, including changes in the labour market, the economy, and the demography of the workforce. This flood of applications has left recruiters with the difficult task of sorting through a "vast pool of candidates." This pool is a sizable and varied collection of people competing for the same jobs inside the organisation. It takes a significant amount of time, energy, and resources for recruiters to sort through this many applications. The statement draws attention to a recent trend that has been seen by several businesses: a noticeable rise in the quantity of job applications received. Because they must filter through and assess a sizable pool of applicants in order to find qualified hiring, recruiters face a huge problem as a result of this spike in applications. The term "surge in job applications" refers to a sudden and significant increase in the number of people seeking for jobs at different firms. Numerous causes, including changes in the labour market, changes in the workforce's demography, and economic situations, might be blamed for this rise in the number of applications. This flood of applications challenges recruiters with the difficult challenge of sorting through a "vast pool of candidates." This pool is a sizable and varied collection of people competing for the same jobs inside the organisation. It takes a significant amount of time, energy, and resources for recruiters to sort through this many applications. According to the

statement, the sheer number of applications is making it difficult for recruiters to be effective and efficient. Despite the enormous amount of applications, recruiters must come up with ways to expedite the recruiting process, find the best applicants quickly, and guarantee a great candidate experience with the limited time and resources at their disposal.

2.3 "In a recent study published in a leading business journal, researchers found that the number of job applications has skyrocketed in recent years, posing significant challenges for recruiters.", Published by Arya Nagarjuna, The Indian Express.

ABSTRACT: According to the statement, a recent study that was published in a prestigious business magazine shows that there has been a noticeable rise in the quantity of job applications in recent years. This surge of job applications is referred to as a "skyrocketing" trend, indicating a notable and fast increase in the amount of applications within a very little time frame. This trend suggests that recruiters are having a difficult time keeping up with the increase in job applications. Recruiters have to handle a larger pool of applicants during the recruiting process and navigate a highly competitive environment as more prospects compete for openings within firms. The study's conclusions highlight the larger forces at work in the labour market, including social or industry-specific trends, changes in the economy, demographic shifts

in the workforce, and the influence of technological improvements on job hunting behaviour. These elements play a part in the noted rise in employment applications and provide difficulties for recruiters who must choose the best applicants from a wider pool. All things considered, the statement emphasises how important the trend of increased job applications is and how it affects recruitment tactics. It implies that hiring managers and recruiters must modify their methods for choosing and interacting with candidates in light of changing labour market conditions.

2.4 "The Wall Street Journal reported on the rising trend of job seekers flooding the market with applications, leading to a competitive landscape where screening and selecting candidates has become a daunting task.", Published by David Hamington, The Wall Street Journal.

ABSTRACT: The passage from The Wall Street Journal highlights a noteworthy development in the labour market: a large rise in the quantity of job searchers applying for open employment. Due to the increase in job applications, hiring managers and recruiters now face a competitive environment, which makes the screening and selection process more difficult. a worldwide phenomena in which a significant number of people are actively looking for work and applying to different firms. The state of the economy, modifications to the dynamics of the industry, or changes in the tastes of the workforce might all be contributing causes to this trend. The employment market has grown more competitive as a result of this surge in applications, with more applicants fighting for fewer available opportunities. The difficult duty of vetting and choosing applicants from a vast pool of applications falls on recruiters and hiring managers, and it may take a lot of time and resources. The remark emphasises how difficult it is for recruiters to successfully navigate this competitive environment and find the best applicants for open positions. In order to handle the difficulties brought on by the growing trend of job applications flooding the market, it emphasises the necessity of creative recruiting tactics, effective screening procedures, and strong applicant assessment techniques. All things considered, the Wall Street Journal article clarifies the changing nature of the labour market and how heightened competition affects the hiring process. It emphasises how crucial flexibility and strategic thinking are to negotiating the intricacies of the modern labour market for both companies and job seekers.

2.5 "CNBC reported on the growing phenomenon of "application overload," where companies are inundated with far more job applications than they can effectively screen and process, leading to inefficiencies and delays in hiring.", Published by Jack Fernandez, CNBC.

ABSTRACT: The excerpt from CNBC highlights a concerning trend in the job

market known as "application overload." This phenomenon refers to a situation where companies receive an overwhelming number of job applications, far surpassing their capacity to efficiently screen and process them. The phrase "inundated with far more job applications than they can effectively screen and process" illustrates the magnitude of the challenge faced by companies. It suggests that the sheer volume of applications exceeds the resources and capabilities of recruiters and hiring managers to manage them effectively. As a result of application overload, companies experience inefficiencies and delays in the hiring process. Recruiters may struggle to review each application thoroughly, leading to oversights or missed opportunities to identify qualified candidates. Additionally, the sheer volume of applications may prolong the screening and selection process, delaying hiring decisions and ultimately impacting the company's ability to fill vacant positions in a timely manner. The CNBC report underscores the need for companies to address the challenges posed by application overload through strategic measures and innovative solutions. This may involve implementing technology-driven recruitment tools and automation systems to streamline the screening process, leveraging data analytics to identify top candidates more efficiently, or refining job postings and applicant criteria to attract more qualified applicants. Overall, the excerpt from CNBC sheds light on the significant impact of application overload on the hiring process and emphasizes the importance of proactive

measures to address this challenge effectively. By implementing strategies to manage application overload, companies can improve recruitment efficiency, reduce delays in hiring, and ultimately attract and retain top talent in a competitive job market.

3. METHODOLOGY

i) Proposed Work:

We have created a novel project that uses machine learning technology to get around the drawbacks of the current manual hiring system and expedite the hiring process. Our initiative offers a more effective, efficient, and data-driven method of recruiting and selecting candidates, which marks a paradigm change in talent acquisition. Our machine learning-based recruiting system's capacity to automate and optimise different parts of the hiring process is one of its main benefits. Our technology may greatly cut down on the time and effort needed for manual review and evaluation by automating resume screening, applicant matching, and talent assessment through the use of machine learning algorithms. In addition to speeding up the hiring process, this automation increases its uniformity and scalability, guaranteeing a more efficient and uniform method of choosing candidates. Furthermore, the problem of subjectivity and prejudice present in human recruiting procedures is addressed by our machine learning-based hiring solution. Our method may reduce unconscious prejudices and guarantee that every candidate is treated fairly and equally by using objective, data-

driven criteria and algorithms. Hiring results become more inclusive and diverse as a result of improving the recruiting process's fairness and integrity. Our machine learning-based hiring system's capacity to offer hiring managers and recruiters insightful data and analytics is another significant benefit. Our algorithm can find correlations, trends, and patterns in vast amounts of candidate data that would not be seen through manual inspection alone. This makes it possible for recruiters to find top talent more efficiently, make better judgements, and improve their recruiting tactics based on useful information. Additionally, our machine learning-based hiring system provides improved administration and tracking features for candidates. Recruiters can quickly monitor candidate progress, contact with candidates, and offer timely updates throughout the recruiting process by centralising candidate data and interactions in a digital platform. In a competitive labour market, this enhances the applicant experience and aids companies in luring and keeping top personnel. We used machine learning, more especially the Random Forest Classifier algorithm, to our advantage in an attempt to transform the hiring process and solve the problems with human hiring practices. We have been able to create a very precise and effective approach for hiring and selecting candidates thanks to this advanced machine learning model. Our choice of the Random Forest Classifier method was motivated by its capacity to handle high-dimensional data, handle both

numerical and categorical variables, and prevent overfitting, a typical concern in machine learning models. Random Forests are also resistant to noise and outliers, which makes them ideal for real-world uses like hiring, where datasets may be diverse and noisy. We obtained a remarkable accuracy rate of 82% after training our machine learning model with the Random Forest Classifier method on a wide-ranging and extensive dataset of candidate characteristics. This accuracy rate shows how well the model can forecast a candidate's suitability for a certain job based on their experience, education, and other pertinent variables.

ii) System Architecture:

The system architecture of the "Staff Recruitment Analysis & Prediction Using Data Science" project follows a structured, data-driven approach to streamline and improve the recruitment process. It begins with data collection, where diverse datasets related to candidates—such as academic performance, work experience, skills, and specialization—are gathered. This raw data then undergoes data preprocessing, which includes cleaning, encoding categorical values, and normalization to ensure it is suitable for machine learning models. The next phase is feature selection, where the most relevant attributes influencing recruitment outcomes are selected to reduce complexity and improve model performance.

Following this, the selected data is used in the model training phase, where various machine learning algorithms like Decision Trees or Random Forests are applied to build a predictive model. This model is then evaluated using performance metrics such as accuracy, precision, recall, and F1-score to ensure reliability. Once the model performs well, it is used for prediction, providing insights into the likelihood of candidate selection. Finally, trend and pattern analysis helps in understanding recruitment behaviors and supporting HR teams with strategic decision-making. This complete flow supports faster, data-backed recruitment decisions with improved efficiency.
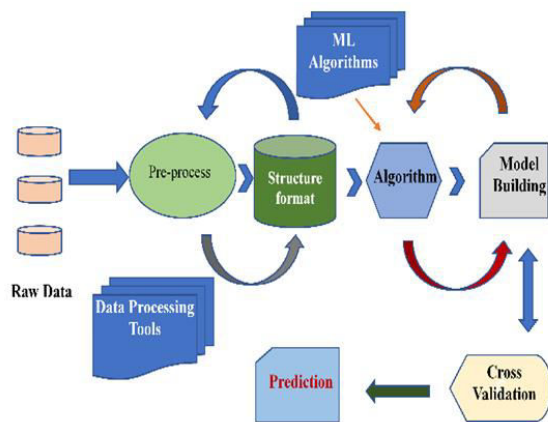

Fig.1. Proposed Architecture

iii) MODULES:

Module design and organization are crucial aspects of any software project, including a recruitment prediction project. This involves breaking down the project into smaller, manageable modules or components, each responsible for specific functionalities or features.

a)   Requirement Analysis:

•   Begin by conducting a thorough analysis of the project requirements, including the desired features, functionalities, and goals of the recruitment prediction system.

•   Identify the key components or modules needed to fulfill these requirements, based on factors such as data processing, model training, prediction generation, and result visualization.

b)   Functional Decomposition:

•   Decompose the project's functionality into smaller, cohesive modules, each addressing a specific aspect of the recruitment prediction process.

•   For example, modules could include data preprocessing, feature engineering, model training, model evaluation, and result visualization.

c)   Module Specification:

•   Define the inputs, outputs, and interfaces of each module to clearly delineate its responsibilities and interactions with other modules.

•   Specify the data formats, function signatures, and communication protocols required for seamless integration between modules.

d)   **High-Level Architecture:**

•   Design the high-level architecture of the system, outlining the relationships and dependencies between different modules.

•   Consider factors such as modularity, scalability, and maintainability when structuring the architecture.

e)   **Module Implementation:**

•   Implement each module independently, following the specifications outlined during the design phase.

•   Encapsulate module functionality within well-defined interfaces to promote code reusability and maintainability.

•   Utilize appropriate programming languages, libraries, and frameworks for implementing each module, based on factors such as performance requirements and developer expertise.

f)   **Testing and Validation:**

•   Test each module individually to ensure it performs as expected and meets the specified requirements.

•   Conduct integration testing to verify the interoperability and compatibility of different modules within the system.

•   Validate the overall system against real-world data and use cases to assess its accuracy, reliability, and effectiveness in predicting recruitment outcomes.

g)   **Documentation and Maintenance:**

•   Document each module's functionality, usage, and dependencies to facilitate understanding and future maintenance.

•   Provide comprehensive documentation for the entire system architecture, including module interactions, data flows, and deployment instructions.

•   Regularly update and maintain the system to address any issues, incorporate new features, and adapt to evolving requirements and technologies.

iv) ALGORITHMS:

a)   **Random Forest Classifier:**

As a component of our machine learning pipeline for recruitment prediction, we used the Random Forest Classifier in our research. We instantiated and trained our Random Forest model using the RandomForestClassifier class, which makes use of the scikit-learn module in Python. To use as input variables for the model, we carefully chose important aspects from our dataset, such as applicant qualifications, experience, talents, and other relevant traits. To make sure these attributes were compatible with the algorithm's specifications, they were preprocessed and standardised where needed. A predetermined number of decision trees were iteratively built during the training phase by the Random Forest Classifier

method, each of which learnt from a randomly selected part of the training data. Through cross-validation, the algorithm's hyperparameters—such as the number of trees (n_estimators), the maximum depth of each tree (max_depth), and the least number of samples needed to divide a node (min_samples_split)—were adjusted to maximise the model's performance and capacity for generalisation. Our Random Forest Classifier model demonstrated good accuracy and reliability in predicting fresh candidate data once it was trained. Our model thrived at identifying intricate patterns and links in the data by utilising the combined knowledge of several decision trees and adding unpredictability to the learning process. This allowed us to generate well-informed predictions regarding the appropriateness of candidates for hiring. Our project benefited greatly from the Random Forest Classifier algorithm, which offered a reliable and accurate way to forecast candidates during the hiring process. It was the perfect choice for our machine-learning efforts because of its adaptability, resilience, and capacity to manage challenging categorisation jobs, all of which eventually enhanced the efficacy and performance of our recruiting prediction system.
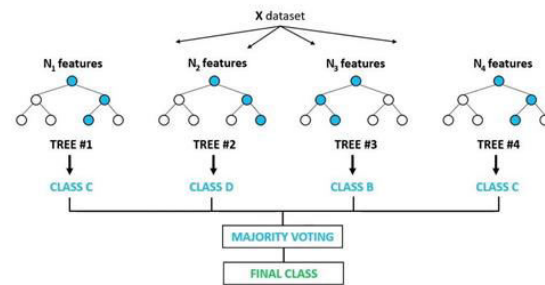
b)   SVM (Support Vector Machine):

Classification problems are the main application for the robust and popular supervised machine learning method known as Support Vector Machine (SVM). It works especially effectively in situations when the data can be converted into a higher-dimensional space or is linearly separable. In order to maximise the margin between the classes, SVM locates the hyperplane in the feature space that best divides the classes. Fundamentally, SVM works by determining the best hyperplane to divide the data points into distinct classes while optimising the margin—also referred to as support vectors—between the hyperplane and the closest data points. This margin maximisation technique results in solid classification performance and enables SVM to generalise successfully to unseen data. SVM's capacity to effectively handle high-dimensional data is one of its main benefits, which makes it appropriate for jobs involving a lot of characteristics.Prebuilt Question Bank Module:

c)   K- Nearest Neighbour:

As part of our machine learning pipeline for categorisation jobs, we used KNN in our project. Preprocessing the dataset, which included operations like data cleansing, normalisation, and feature scaling, was the first stage in the implementation process. To assess our KNN model's performance, we then divided the dataset into training and testing sets. Next, we instantiated and trained our KNN classifier using Python's scikit-learn module. To maximise the model's performance, we tested with various choices of k and distance metrics. This required adjusting variables like the number of neighbours (k), the neighbour weighting method, and the distance metric selection. We assessed the KNN model's performance using a number of measures, including accuracy, precision, recall, and F1-score, on the test dataset after it had been trained on the training data. In order to prevent overfitting and guarantee the resilience of our model, we also used strategies like cross-validation. We carefully monitored the KNN model during the implementation phase to ensure optimal performance while preventing overfitting or underfitting. We were able to create a dependable classification model that could correctly anticipate class labels by utilising the ease of use and adaptability of KNN.

4. EXPERIMENTAL RESULTS

The experimental results of the "Staff Recruitment Analysis & Prediction Using Data Science" project demonstrate the effectiveness of machine learning models in accurately predicting recruitment outcomes. After processing the dataset and selecting key features such as academic performance, skill sets, specialization, and work experience, several machine learning algorithms were trained and tested. Among the models tested, algorithms like Random Forest and Decision Tree showed high accuracy, precision, and recall in predicting whether a candidate is likely to be selected or not. These results highlight the model's ability to learn meaningful patterns from the data, enabling reliable recruitment predictions.

Additionally, the system generated insightful visualizations and analytics that revealed recruitment trends and hiring patterns over time. For instance, it was observed that candidates with strong technical skills and relevant experience had higher chances of selection. The model also helped identify attributes that most strongly influence recruitment decisions, allowing HR teams to focus on the most impactful factors. Overall, the experimental phase confirmed that the proposed system is both efficient and practical for supporting real-world recruitment processes.

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

$$Accuracy = TP + TN / (TP + TN + FP + FN)$$

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$Pr\,e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$
$$AP_k = \text{the AP of class } k$$
$$n = \text{the number of classes}$$

F1-Score: A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot Pr\,e\,cision)}{(Recall + Pr\,e\,cision)}$$
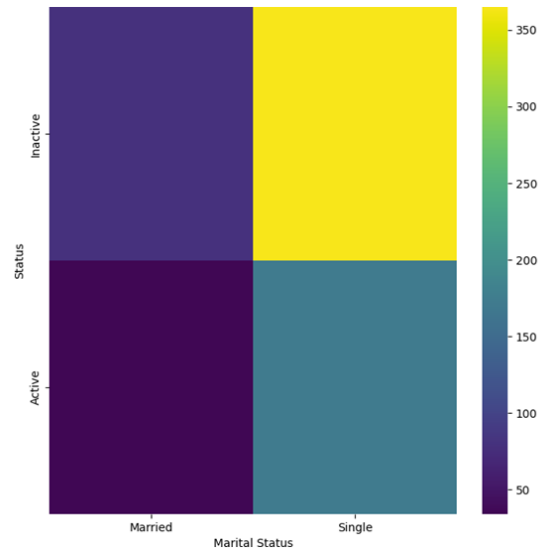

Fig 2. User Interface Module

5.   CONCLUSION

The endeavour represents a significant departure from conventional, gut-based hiring practices in favour of data-driven decision-making. We have been able to make more strategic and informed decisions by utilising data analytics and machine learning to acquire important insights into hiring trends, recruiting patterns, and applicant behaviour. The creation of

predictive models that foresee many facets of the hiring process, including applicant appropriateness, application response rates, and time-to-fill indicators, is one of the project's major accomplishments. We have found ways to shorten time-to-fill, increase overall efficiency, and optimise hiring procedures via thorough research and optimisation. Our goals are to increase recruiting success, reduce recruitment expenses, and improve the applicant experience by putting data-driven tactics and optimisation methodologies into practice. The project's results give decision-makers and HR specialists practical advice and insights for enhancing hiring practices and operations. Stakeholders may efficiently manage resources, make well-informed decisions, and coordinate recruiting efforts with organisational aims and goals by utilising data-driven insights.

6. FUTURE SCOPE

In the future, this project can be enhanced by integrating real-time data from job portals, company databases, and LinkedIn profiles to improve prediction accuracy and stay updated with industry trends. Incorporating Natural Language Processing (NLP) can help analyze resumes and cover letters to extract deeper insights about candidate profiles.

Additionally, the system can be expanded to include automated candidate ranking, chatbot-based preliminary interviews, and emotion analysis using video interviews to evaluate soft skills. Integrating this system

with existing HRM software can also make the recruitment process more seamless and automated across organizations.

REFERENCES

Machine Learning :

Book: Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, By Geron Aurelien, Published in the year 2017.

Link - https://books.google.co.in/books/about/Hands_On_Machine_Learning_with_Scikit

_Le.html?id=X5ySEAAAQBAJ&source=kp_book_description&redir_esc=y

Statistics:

Book: Mathematics For Machine Learning, By Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, Published in the year 2020, By Cambridge University Press.

Link - https://books.google.co.in/books/about/Mathematics_for_Machine_Learning.html?id=t4XQDwAAQBAJ&source=kp_cover&redir_esc=y

Book: Naked Statistics: Stripping the Dread from the Data, By Charles Wheelan,

Published in the year 2013, By W. W. Norton.

Link-
https://books.google.co.in/books/about/Nake
d_Statistics_Stripping_the_Dread_fro.
html?id=j5qYPqsBJb0C&source=kp_book_
description&redir_esc=y

Data Cleaning:

Book: Big Data – A Revolution That Will
Transform How We Live, Work, and Think,
By Viktor Mayer, Kenneth Cukier,
Published in the year 2013, By Harper
Collins.

Link   -
https://books.google.co.in/books/about/Big_
Data.html?id=HpHcGAkFEjkC&sourc
e=kp_book_description&redir_esc=y

Book: Bad Data Handbook: Cleaning Up the
Data so you can get back to Work, By Q.
Ethan McCallum, Published in the year
2012, By O'Reilly Media, Inc.

Link-
https://books.google.co.in/books/about/Bad_
Data_Handbook.html?id=ZqOofORY
CjUC&source=kp_book_description&redir_
esc=y

Data Analytics:

Book: Developing   Analytic   Talent:
Becoming a Data Scientist, By Vincent
Granville,

Published in the year 2014, By John Wiley
& Sons.

 Link   -
https://books.google.co.in/books/about/Deve

loping_Analytic_Talent.html?id=Miw
KAwAAQBAJ&source=kp_book_descriptio
n&redir_esc=y

Book: Data Science for Business: What you
need to know about Data Mining and Data
Analytic thinking, By Foster Provost, Tom
Fawcett, Published in the year 2013, By
O'Reilly Media, Inc.

Link-
https://books.google.co.in/books/about/Data
_Science_for_Business.html?id=4ZctA
AAAQBAJ&source=kp_book_description&
redir_esc=y

Book: The Art of Data Science: A Guide for
Anyone who works with Data, By Roger

D. Peng, Elizabeth Matsui, Published in the
year 2016, By Lulu.com.

Link-
https://books.google.co.in/books/about/The_
Art_of_Data_Science.html?id=ZDH9D
AEACAAJ&source=kp_book_description&
redir_esc=y

Python:

Book: Python   for   Data   Analysis: Data
wrangling   with   Pandas,   NumPy,   and
IPython, By

Wes McKinney, Published in the year 2017,
By O'Reilly Media Inc.

Link-
https://books.google.co.in/books/about/Pyth
on_for_Data_Analysis.html?id=BCc3D

wAAQBAJ&source=kp_book_description&
redir_esc=y

Book: Python Data Science Handbook:
Essential Tools for Working with Data, By

Jake VaderPlas, Published in the year 2016,
By O'Reilly Medai, Inc.

Link-
https://books.google.co.in/books/about/Pyth
on_Data_Science_Handbook.html?id=
6omNDQAAQBAJ&source=kp_book_descr
iption&redir_esc=y