CSCE 5215: MACHINE LEARNING

Project Increment 2

# Project Description:

1. **Project Title and Team Members**

Project Title: Strategic Employee Turnover Prediction System

Team Members:

* Sai Kishore Reddy Bobbili
* Lokesh Bathula
* Bhargavi Guda
* Venkata Bhargavi Nalamothu

1. **Goals and Objectives:**

**Motivation:**

The motivation behind using reliable machine learning approaches for predicting employee turnover is to help organizations identify and understand factors that contribute to employees leaving their jobs. By leveraging machine learning, we aim to develop accurate models that can foresee turnover risks. This is crucial for companies of all sizes as it enables proactive measures to retain valuable employees, ultimately fostering a more stable and productive work environment. Using trustworthy machine learning techniques ensures precise insights into employee behaviour, aiding in the development of effective retention strategies and overall workforce management.

**Significance:**

Employee turnover prediction is significant as it enables the organizations to make informed decisions that impact financial stability, workforce planning, employee satisfaction and overall organizational success. It represents a proactive approach to talent management, ensuring a more stable, engaged, and productive workforce. There are some key aspects that underscore the significance of employee turnover prediction like cost implications, strategic workforce planning, employee engagement and satisfaction, maintaining productivity, succession planning and cultural stability**.**

**Objectives:**

The objectives of employing reliable machine learning approaches for employee turnover prediction encompass identifying key factors influencing turnover, constructing accurate predictive models based on historical data, tailoring t argeted retention strategies, facilitating data-driven decision-making for workforce management, optimizing resource allocation by focusing on high-risk areas, establishing a continuous improvement feedback loop for model refinement, and ensuring ethical and fair application. These objectives collectively aim to empower organizations to proactively address turnover risks, enhance the overall work environment, and foster effective employee retention strategies through the informed utilization of machine learning insights.

**Features:**

Employee turnover prediction involves several key features that collectively contribute to its effectiveness in helping organizations anticipate and manage workforce changes. Firstly, data analysis plays a pivotal role, utilizing historical and current employee data to identify patterns and trends indicative of potential turnover. Machine learning algorithms and statistical models are often employed to make accurate predictions based on various factors such as job satisfaction, performance, and tenure. Furthermore, the incorporation of qualitative data, including employee feedback and sentiments, enhances the predictive accuracy by capturing nuanced aspects of job dissatisfaction. Overall, employee turnover prediction features a multidimensional approach that blends quantitative and qualitative insights to provide actionable information for strategic workforce planning and retention efforts.

**Increment 2:**

* Related Work (Background)

**Dataset:**

The provided dataset appears to be a collection of employee-related information, likely from a human resources or workforce management system. Each row represents an employee's data, and the columns contain various attributes such as satisfaction level, last evaluation score, number of projects, average monthly hours worked, time spent at the company, work accident status, promotion within the last five years, and the employee's department and salary level. These attributes provide a comprehensive overview of the employees' performance, engagement, and general work-related information. The dataset seems suitable for analysis related to employee satisfaction, performance evaluation, and other workforce-related insights. Additionally, the inclusion of department and salary information allows for the examination of trends or patterns within specific organizational units and income levels.

**Detail design of Methods:**

A screenshot of a computer screen

Description automatically generated

**Analysis:**

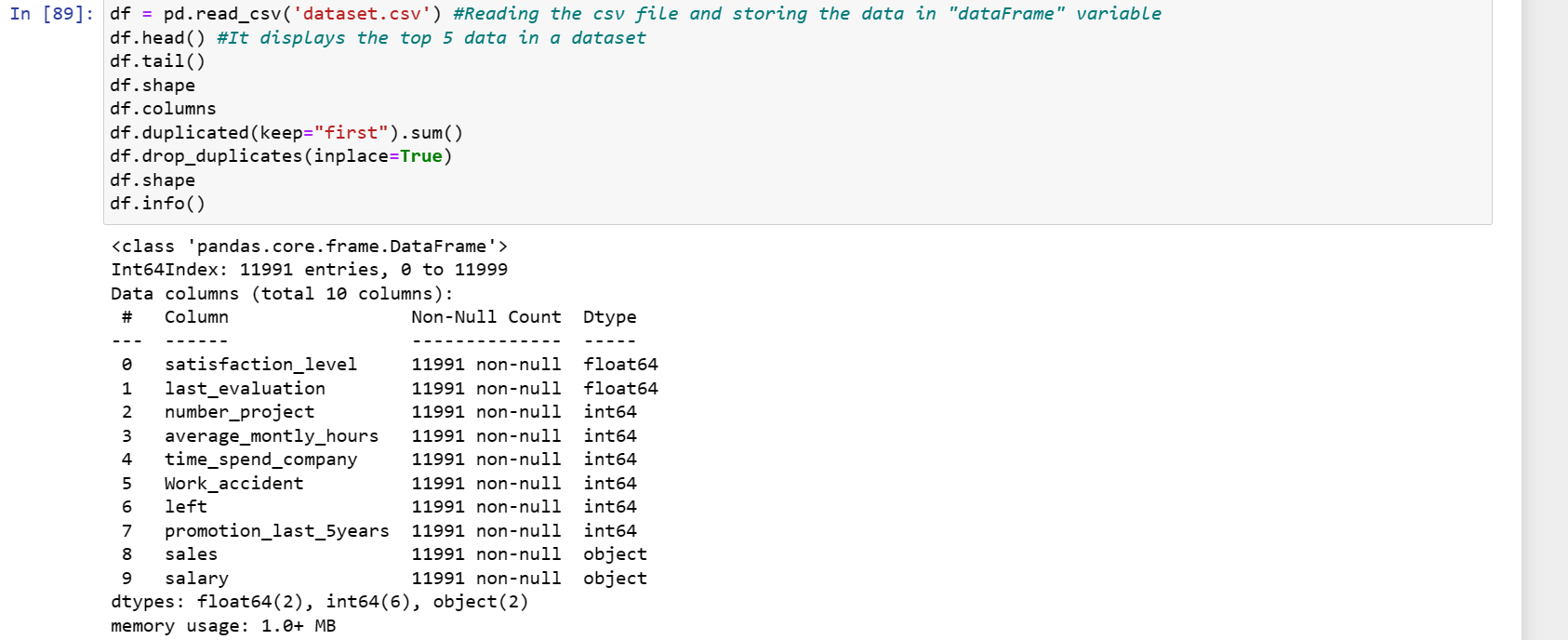
This analysis serves as a practical guide for organizations looking to grasp the complexities of employee turnover. It helps companies align their strategies with data-driven insights, fostering a workforce that is resilient and content. In essence, it provides a roadmap for companies to navigate the challenges of managing turnover effectively.

**Implementation:**

Model creation and testing:

Data Preprocessing

Manipulating all the unwanted data to meaningful data.

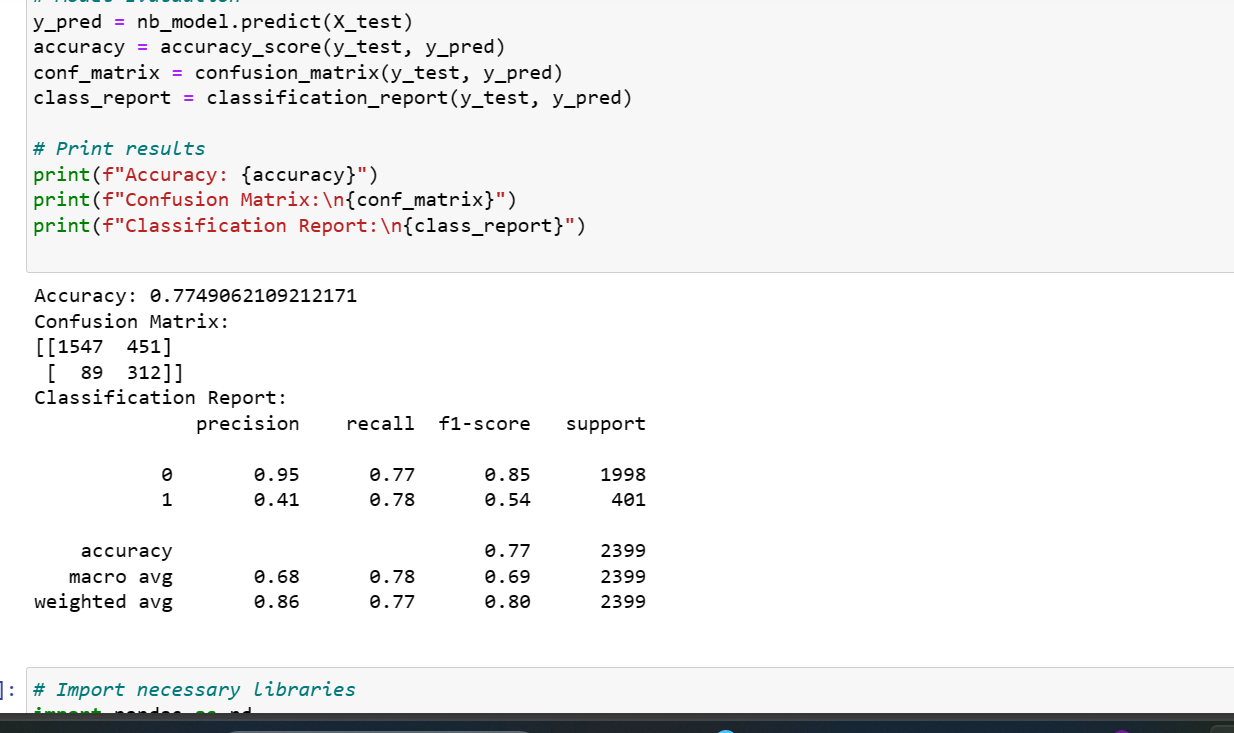
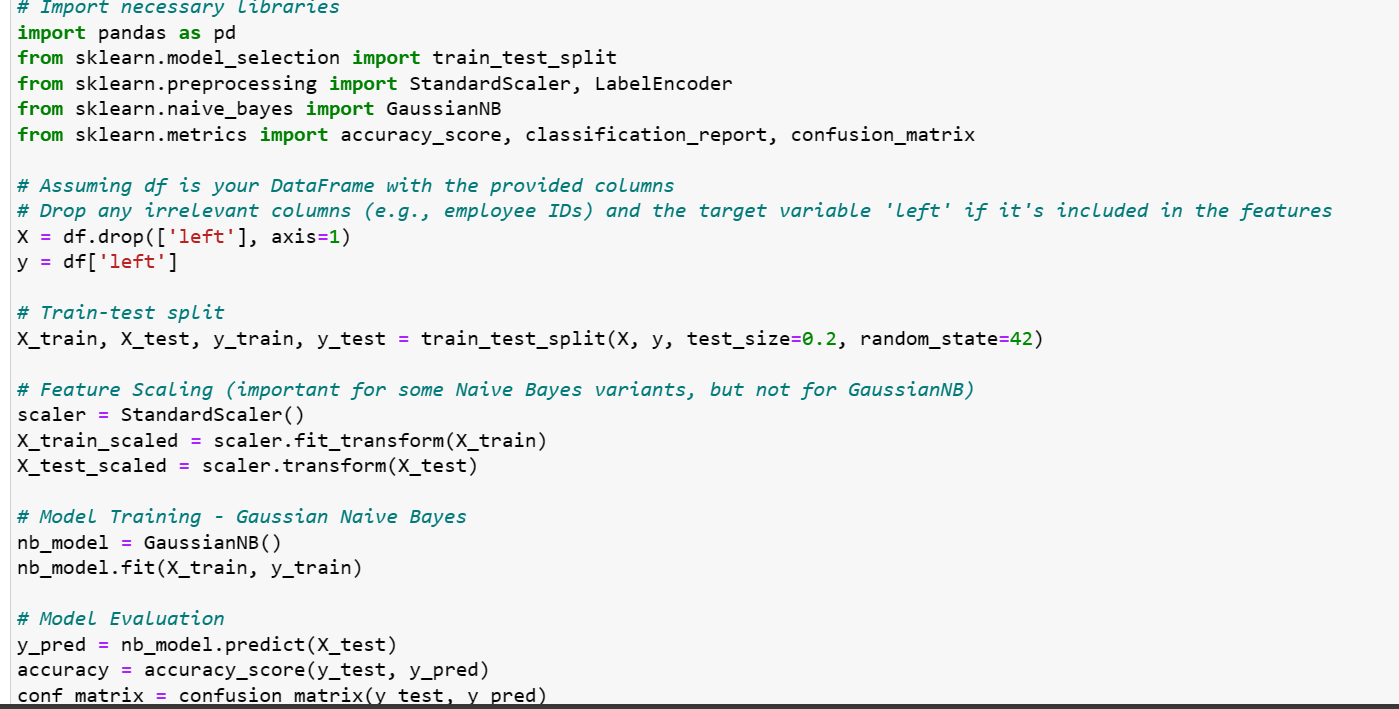


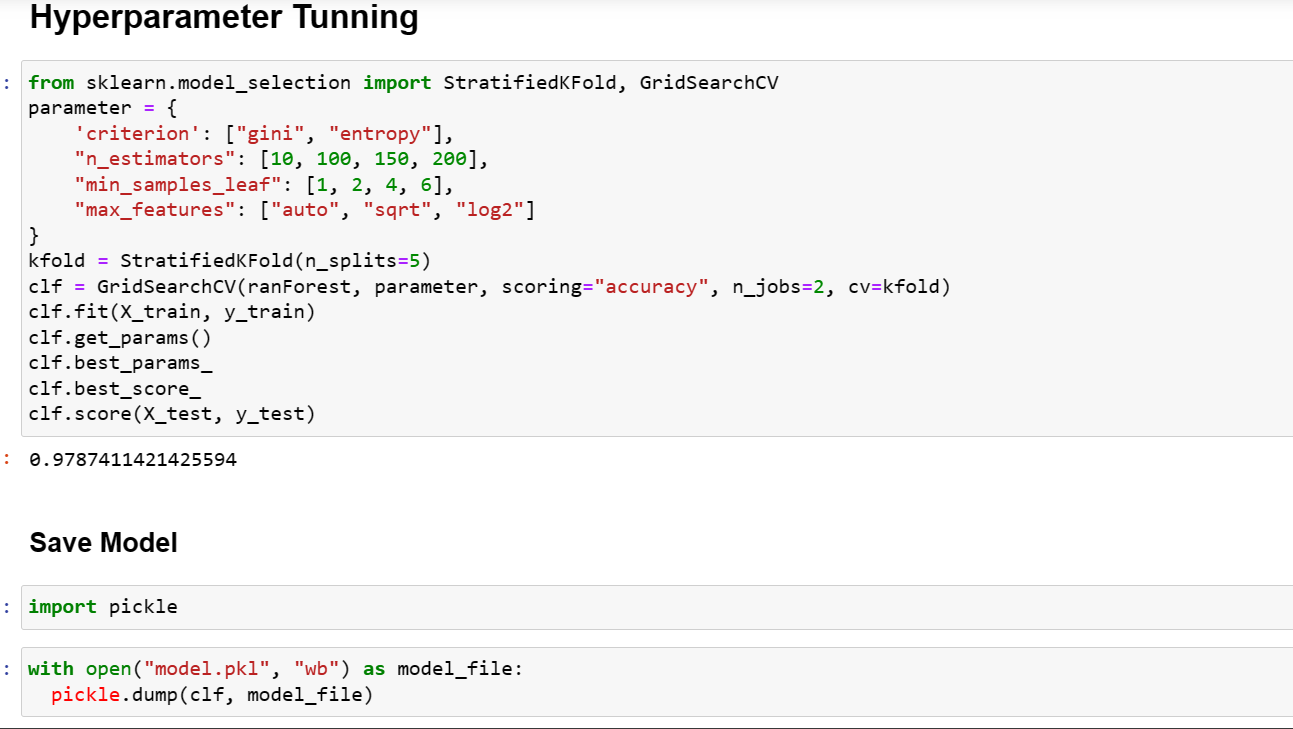
## Exploratory Data Analysis:

## 

**Algorithms**

Algorithms model that are used to predict and find the accuracy for the employee to leave an organization.





**Performance evolution:**

**Accuracy:**

|  |  |
| --- | --- |
| **Algorithms** | **Accuracy** |
| **Random Forest Classifier** | **0.98** |
| Logistic Regression | 0.84 |
| Support Vector Machine | 0.83 |
| K-Nearest Neighbour | 0.95 |
| Naïve Bayes | 0.77 |

**Precision:**

|  |  |
| --- | --- |
| **Algorithms** | **Precision** |
| Random Forest Classifier | 0.98 |
| Logistic Regression | 0.86 |
| Support Vector Machine | 0.83 |
| K-Nearest Neighbour | 0.97 |
| Naïve Bayes | 0.95 |

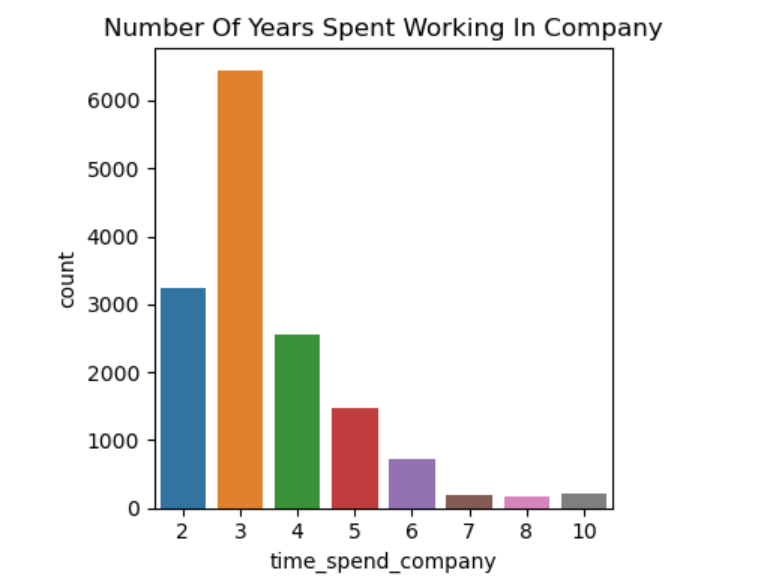
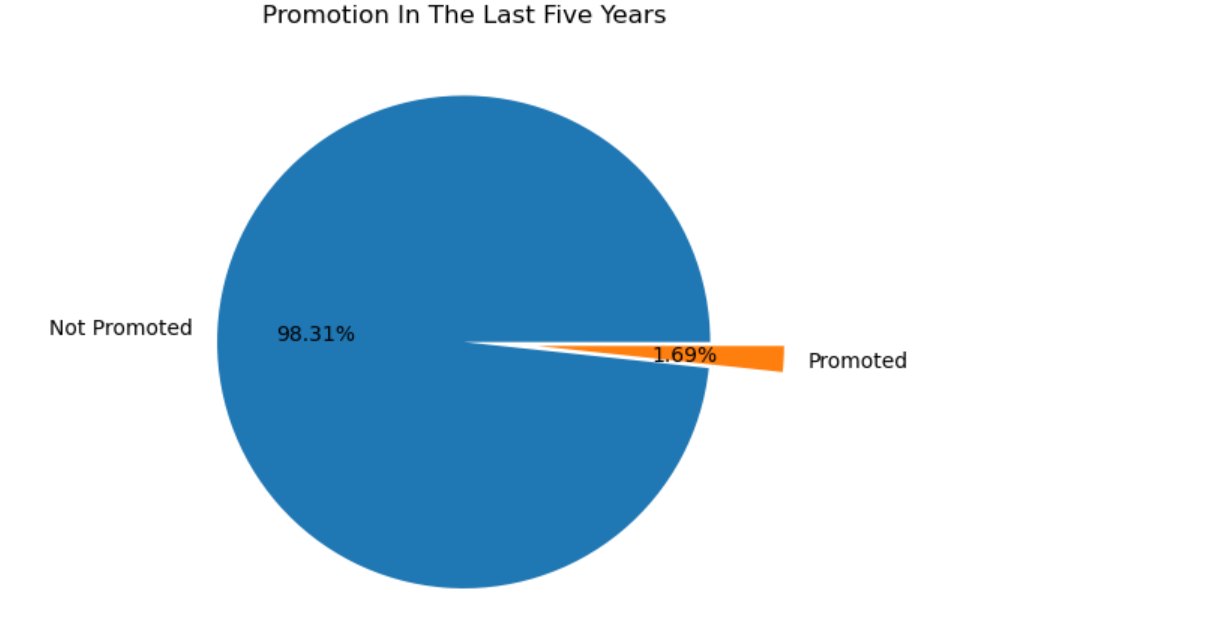
**Recall:**

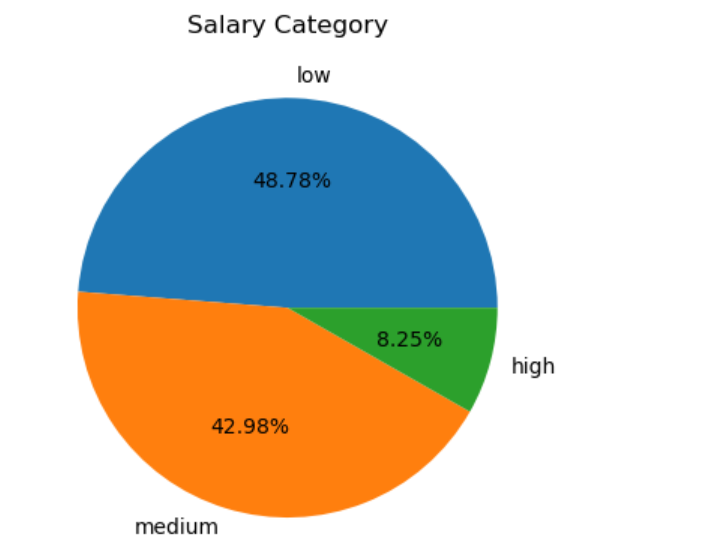
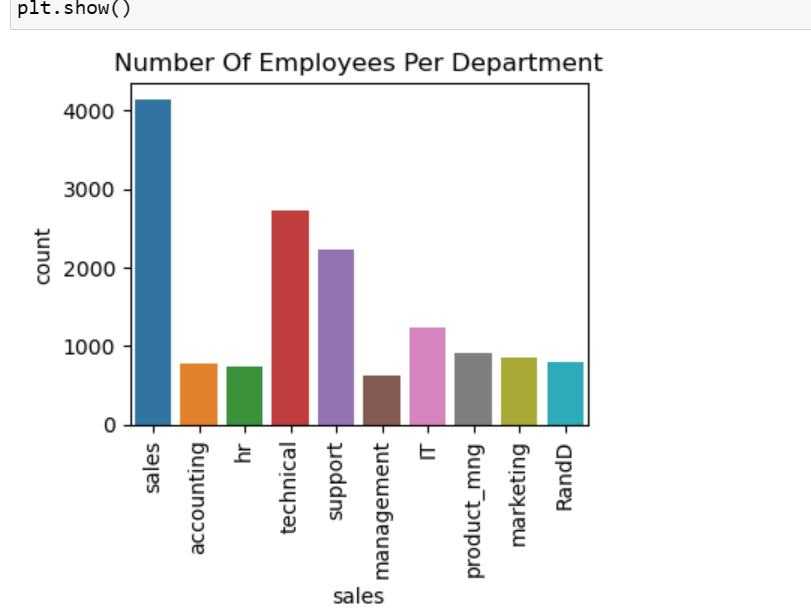
|  |  |
| --- | --- |
| **Algorithms** | **Recall** |
| Random Forest Classifier | 1.00 |
| Logistic Regression | 0.96 |
| Support Vector Machine | 1.00 |
| K-Nearest Neighbour | 0.97 |
| Naïve Bayes | 0.77 |

**F1 Score:**

|  |  |
| --- | --- |
| **Algorithms** | **F1-Score** |
| Random Forest Classifier | 0.99 |
| Logistic Regression | 0.91 |
| Support Vector Machine | 0.91 |
| K-Nearest Neighbour | 0.97 |
| Naïve Bayes | 0.85 |

**Premilary results:**





**Project Management**

**Implementation report**

**Work completed:**

* Data Preprocessing and Exploratory Data Analysis: The dataset underwent thorough cleaning and preprocessing to ensure data quality.
* Exploratory Data Analysis (EDA) was conducted to gain insights into the dataset's characteristics, providing a foundational understanding.
* Algorithm Testing, Selection, and Hyperparameter Tuning: We tested six machine learning algorithms, namely Logistic Regression, Random Forest Classifier, K-Nearest Neighbors, Decision Tree Classifier, Gaussian Naive Bayes, and Support Vector Classifier.The performance of each algorithm was evaluated using various metrics to identify the most effective model for our task.
* Hyperparameter tuning was implemented to fine-tune the selected model for optimal performance.
* Model Fitting and Evaluation:The selected model was fitted with the preprocessed dataset.
* Model evaluation involved assessing accuracy, precision, recall, and F1-score to measure its overall performance.
* This comprehensive approach ensures that our models not only underwent rigorous testing and selection but also benefited from hyperparameter tuning to enhance their predictive capabilities. The evaluation metrics provide a clear picture of how well our model performs in predicting the desired outcome.

**Responsibility**

* Worked with Logistic Regression Model, EDA (Exploratory Data Analysis) for monthly hours column 🡪Sai Kishore Reddy Bobbili
* Worked with K-Nearest Neigbor model and EDA for no.of projects column

**🡪** Lokesh Bathula

* Worked with Naïve Bayes Model and EDA for the Promotion Made In The Last 5 Years column 🡪 Bhargavi Guda
* Worked with Random Forest and EDA for the work accidents column

🡪 Venkata Bhargavi Nalamothu

**Contributions**

* Sai Kishore Reddy Bobbili🡪25%
* Lokesh Bathula 🡪25%
* Bhargavi Guda🡪25%
* Venkata Bhargavi Nalamothu🡪25%

**References**

1. Self T T, Jolly P M, Gordon S E. Family-supportive supervisor behaviors and employee turnover intention in the foodservice industry: does gender matter? [J]. International Journal of Contemporary Hospitality Management, 2022, 34(3): 1084–1105.<https://www.emerald.com/insight/content/doi/10.1108/IJCHM-02-2021-0280/full/html>.
2. Peltokorpi V, Allen D G, Shipp A J. Time to leave? The interaction of temporal focus and turnover intentions in explaining voluntary turnover behaviour[J]. Applied Psychology, 2023, 72(1): 297–316.<https://iaap-journals.onlinelibrary.wiley.com/doi/10.1111/apps.12378>.
3. Szajna A, Kostrzewski M. AR-AI tools as a response to high employee turnover and shortages in manufacturing during regular, pandemic, and war times[J]. Sustainability, 2022, 14(11): 6729.
4. B. Holtom, T. Mitchell, T. Lee, and M. Eberly, “Turnover and retention research: A glance at the past, a closer review of the present, and a venture into the future”, Academy of Management Annals, 2: 231-274, 2008. <https://www.researchgate.net/publication/233055591_5Turnover_and_Retention_Research_A_Glance_at_the_Past_a_Closer_Review_of_the_Present_and_a_Venture_into_the_Future>.
5. Alduayj, S.S., Rajpoot, K.: Predicting employee attrition using machine learning. In: IIT, pp. 93–98. IEEE (2018).<https://dergipark.org.tr/en/pub/cukurovaumfd/issue/67373/1040487>.
6. Fallucchi, F., Coladangelo, M., Giuliano, R., Luca, E.W.D.: Predicting employee attrition using machine learning techniques. Computer **9**(4), 86 (2020).<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9251085/>.
7. Griffeth, R.W., Hom, P.W., Gaertner, S.: A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. J. Manag. 26(3), 463–488 (2000)<https://www.scirp.org/(S(vtj3fa45qm1ean45vvffcz55))/reference/ReferencesPapers.aspx?ReferenceID=1909452>
8. Hassan, S.: The importance of role clarification in workgroups: Effects on perceived role clarity, work satisfaction, and turnover rates. Public Adm. Rev. 73(5), 716–725 (2013).<https://www.ctg.albany.edu/media/pubs/pdfs/clarityofroles.pdf>.
9. Hom, P., Lee, T., Shaw, J., Hausknecht, J.: One hundred years of employee turnover theory and research. J. Appl. Psychol. 102, 530 (2017).<https://psycnet.apa.org/doiLanding?doi=10.1037%2Fapl0000103>.
10. Tanova, C., Holtom, B.C.: Using job embeddedness factors to explain voluntary turnover in four European countries. Int. J. Human Res. Manag. **19**, 1553–1568 (2008).<https://www.tandfonline.com/doi/abs/10.1080/09585190802294820>.
11. Ngo-Henha, P.E.: A review of existing turnover intention theories. Int J Econ. Manag. Eng. 11, 2760–2767 (2017)<https://publications.waset.org/10008789/a-review-of-existing-turnover-intention-theories>.

**Deployment**

**GitHub Link:** [**Click Here**](https://github.com/Kishore15250/Strategic-Employee-Turnover-Prediction-Using-Machine-Learning.git)

Note: The references must be linked in the document.

# Submission Guidelines:

1. Submit your source code and documentation to GitHub and Canvas.
2. The GitHub link should be provided in the document.
3. Source code has to be properly commented.
4. The documentation should include the visualization of preliminary results.
5. Submit a brief demo video showing your source code and project tasks with a voice over explaining your work through the submission link.
6. The similarity score for the document should be less than 15%.