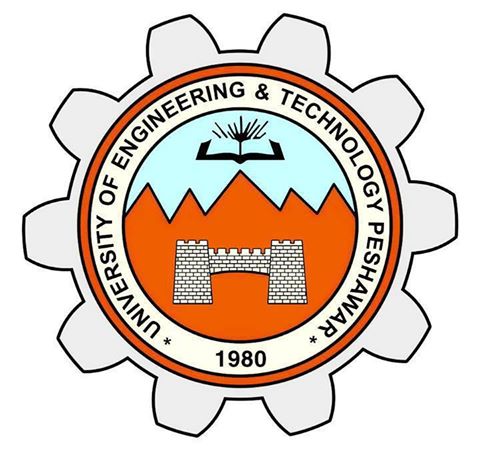
**Predicting Employees Attrition and Identify Factors Leading to Attrition**



**M.Sc Mini Project Proposal**

For

M.Sc in Computer Science

Department of Computer Science and Information Technology

**Submitted By**

Muhammad Waqas (F20PWMCS0138)

**Supervisor**

Dr. Iftikhar Ahmad

University of Engineering & Technology, Peshawar, Pakistan

|  |  |
| --- | --- |
| UET new Logo | **University of Engineering and Technology,**  **PEshawar, Pakistan**  M.Sc MINI pROJECT Proposal |

|  |  |
| --- | --- |
| Proposal Title | **Predicting Employees Attrition and Identify Factors Leading to Attrition** |
| Department | Department of Computer Science and Information Technology |
| Student Name | Muhammad Waqas |
| Father’s Name | Muhammad Nabi |
| Contact No. | 03059987768 |
| Email | mwaqas@uetpeshawar.edu.pk |
| Registration No. | F20PWMCS0138 |
| Date of Registration in the Program | October 2020 |
| Supervisor | Dr. Iftikhar Ahmad |

**Courses Studied**

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Course No. and Name** | **Grade** |
| 1. | Advance Theory of Computation (core) |  |
| 2. | Software Project Management |  |
| 3. | Computer Security |  |
| 4. | Advance Analysis of Algorithm (core) |  |
| 5. | Advance Operation Research |  |
| 6. | Principles of Data Science |  |
| 7. | Machine Learning |  |
| 8. | Research Methodology (core) |  |
| 9. | Web Security |  |
| **CGPA** | |  |

**Abstract:** Employee attrition is a major issue for businesses since it affects productivity and profitability. Predicting employee turnover and recognizing its root causes can assist organizations in taking proactive steps to retain staff and maintain a healthy work environment. The goal of this project is to create a machine learning model that can effectively forecast staff turnover based on parameters such as job satisfaction, salary, work-life balance, and so on. I will also do exploratory data analysis to discover the major causes of attrition and provide organizations with actionable recommendations.

1. **Introduction**

Employee attrition is a common problem faced by organizations across various industries. Losing skilled and experienced employees can have a significant impact on an organization's productivity, profitability, and overall success. Furthermore, recruiting and training new employees can be time-consuming and expensive. As a result, it is critical for organizations to identify the leading causes of attrition and take proactive measures to retain their employees.

Machine learning models may effectively forecast whether an employee is likely to leave the organization by analyzing several parameters such as job satisfaction, salary, work-life balance, and so on. Exploratory data analysis can also assist discover the major reasons for attrition in addition to predicting it.

My goal in this project is to create a machine learning model that can accurately predict employee attrition based on a variety of parameters. I will also conduct exploratory data analysis to discover the major causes of attrition and provide organizations with actionable insights. This project's findings can assist organizations in taking proactive actions to retain staff and maintain a healthy work environment.

1. **The Problem Statement**

Employee attrition is a major challenge for organizations, as it affects their productivity and profitability. While there are various factors that contribute to attrition, organizations often struggle to identify the leading causes and take practical measures to retain their employees. Therefore, there is a need for a predictive model that can accurately predict employee attrition based on various factors such as job satisfaction, salary, work-life balance, etc. In addition, there is a need for exploratory data analysis to identify the leading causes of attrition and provide actionable insights to organizations. This project aims to address these needs by developing a machine learning model that can predict employee attrition and identify its leading causes. The insights gained from this project can help organizations take proactive measures to retain their employees and maintain a healthy work environment.

1. **Aims and Objectives**

The aim of this project is to develop a machine learning model that can accurately predict employee attrition and identify its leading causes.

* Collect and preprocess employee data from the organization's HR system.
* Perform exploratory data analysis to identify the leading causes of attrition.
* Develop a machine learning model to predict employee attrition based on various factors such as job satisfaction, salary, work-life balance, etc.
* Evaluate the performance of the machine learning model using appropriate metrics such as accuracy, precision, recall, etc.

1. **Literature Review**

Employee attrition is a complex problem that has been studied extensively in the literature. Various studies have identified different factors that contribute to attrition, such as job satisfaction, salary, work-life balance, organizational culture, and job security. For instance, a study by Bhatnagar (2012) found that job satisfaction is the most significant predictor of employee turnover. Similarly, a study by Koustelios and Bagiatis (2014) identified salary as a key factor that influences employee retention.

In recent years, machine learning has emerged as a powerful tool for predicting employee attrition. Several studies have explored the use of machine learning algorithms such as logistic regression, decision trees, and random forests for predicting attrition. For instance, a study by Gupta and Sharma (2018) used a random forest algorithm to predict employee attrition based on various factors such as age, gender, job level, etc. The study reported an accuracy of 87% in predicting attrition.

Exploratory data analysis is another important tool for identifying the leading causes of attrition. By analyzing the data, organizations can gain insights into the factors that contribute to attrition and take proactive measures to address them. For instance, a study by Ramlall (2004) used exploratory data analysis to identify the leading causes of attrition in a retail organization. The study found that low job satisfaction, lack of recognition, and poor work-life balance were the main factors contributing to attrition.

Overall, the literature suggests that employee attrition is a complex problem that requires a multifaceted approach. Machine learning can be used to predict attrition based on various factors, while exploratory data analysis can help identify the leading causes of attrition. By combining these techniques, organizations can gain valuable insights into their workforce and take proactive measures to retain their employees.

1. **Methodology**

This project will be implemented using Jupyter Notebook, Python, and various Python libraries such as Pandas, NumPy,s Scikit-learn, and Plotly. The project will follow the following steps:

* Data Collection: The first step will be to collect employee data from the organization's HR system. The data will be collected in a CSV format and stored in a local directory.
* Data Preprocessing: The collected data will be preprocessed to remove any missing values, duplicates, or outliers. The data will also be transformed and normalized as per the requirements of the machine learning algorithm.
* Exploratory Data Analysis: Exploratory data analysis (EDA) will be performed to gain insights into the relationships between various factors and employee attrition. This will involve visualizing the data using various plots and charts using Plotly.
* Feature Selection: Based on the insights gained from EDA, relevant features will be selected for training the machine learning model. This will involve using techniques such as correlation analysis and feature importance analysis.
* Model Development: A machine learning model will be developed using the Random Forest algorithm. The model will be trained on the selected features using Scikit-learn library.
* Model Evaluation: The performance of the machine learning model will be evaluated using appropriate metrics such as accuracy, precision, recall, etc. Cross-validation techniques such as k-fold cross-validation will also be used to ensure that the model is not overfitting.
* Results Visualization: The results of the analysis will be visualized using various plots and charts using Plotly.

Overall, this methodology aims to provide a comprehensive approach to predicting employee attrition and identifying its leading causes. By using Jupyter Notebook, Python, and various Python libraries such as Pandas, NumPy, Scikit-learn, and Plotly, we aim to develop a robust and accurate machine learning model that can help organizations take proactive measures to retain their employees.

1. **Data Source**

The Data will be taken from the following source:

Source: <https://www.kaggle.com/datasets/thedevastator/employee-attrition-and-factors>

# References:

1. Bhatnagar, J. (2012). Turnover intentions and its predictors: A study of Indian IT professionals. Journal of Business and Psychology, 27(2), 225-242.
2. Gupta, A., & Sharma, R. K. (2018). Employee turnover prediction using machine learning: A case study. International Journal of Advanced Computer Science and Applications, 9(8), 329-334.
3. Koustelios, A., & Bagiatis, V. (2014). Factors affecting employee retention: Evidence from the Greek shipping industry. Maritime Policy & Management, 41(1), 47-65.
4. Ramlall, S. (2004). A review of employee motivation theories and their implications for employee retention within organizations. Journal of American Academy of Business, Cambridge, 5(1/2), 52-63.