Project Charter

Employee Attrition Predictive Analytics

Prepared for: Human Resources Department, FACE Prep Campus

Prepared by: Data Science Capstone Team
([Student's Name(s)])

Project Mentor: [Instructor's Name / Your Name as Mentor]

Version: 1.0

Date: May 9, 2025

Abstract

This document outlines the charter for the "Employee Attrition Predictive Analytics" project undertaken by the Data Science Capstone Team for FACE Prep Campus. The primary objective is to develop a robust predictive model and an interactive dashboard to identify employees at high risk of attrition. This initiative aims to provide the Human Resources department with actionable insights, enabling proactive retention strategies and ultimately reducing turnover-related costs and preserving institutional knowledge. The project will leverage historical employee data and utilize Python-based data science tools, including Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, and Streamlit for the GUI.

Contents

1 Introduction and Background

FACE Prep Campus (hereinafter "FACE Prep") recognizes that employee attrition is a critical business challenge, leading to significant financial costs associated with recruitment, onboarding, and lost productivity, as well as intangible costs such as diminished team morale and loss of valuable institutional knowledge. The Human Resources (HR) department currently employs traditional methods for monitoring employee engagement and addressing turnover, which are often reactive. This project proposes a data-driven, proactive approach to identify and mitigate attrition risks.

2 Problem Statement

The inability to proactively identify employees at high risk of attrition limits FACE Prep's capacity to implement timely and targeted retention strategies. This results in avoidable employee turnover, impacting operational efficiency and strategic objectives. A predictive analytics solution is required to provide early warnings and insights into the key drivers of attrition.

3 Project Goals and Objectives

The overarching goal of this project is to empower FACE Prep's HR department with a data-driven tool to reduce employee attrition.

3.1 Primary Objectives

- **Develop a Predictive Model:** Construct and validate a machine learning model using Scikit-learn to accurately predict the likelihood of an employee leaving FACE Prep.
- **Identify Key Attrition Drivers:** Utilize statistical analysis and model interpretability techniques to uncover the most significant factors contributing to employee attrition.
- **Develop an Interactive Dashboard:** Create a user-friendly GUI using Streamlit to:
  - Allow HR personnel to input employee characteristics and receive real-time attrition risk predictions.
  - Visualize key attrition trends and model insights.
- **Provide Actionable Insights:** Deliver a comprehensive report and presentation summarizing findings, model performance, and data-backed recommendations for potential intervention areas.
- **Knowledge Transfer:** Ensure adequate documentation and, if applicable, a brief training session for HR personnel on using the developed dashboard.

4 Scope of Work

4.1 In-Scope Activities

- Data Acquisition and Understanding: Utilizing the "IBM HR Analytics Employee Attrition Performance" dataset.
- Data Cleaning and Preprocessing: Handling missing values, outliers, data type conversions, and encoding.

- Exploratory Data Analysis (EDA): In-depth statistical and visual analysis of variables and their relationship with attrition.
- Feature Engineering and Selection: Creating new informative features and selecting the most relevant ones for modeling.
- Model Development and Training: Experimenting with at least two Scikit-learn classification algorithms (e.g., Logistic Regression, Random Forest).
- Model Evaluation and Validation: Rigorous assessment using appropriate metrics (Accuracy, Precision, Recall, F1-score, ROC-AUC), focusing on the "Attrition: Yes" class.
- Feature Importance Analysis: Identifying and ranking influential predictors.
- GUI Development: Building an interactive dashboard using Streamlit.
- Documentation: Comprehensive Jupyter Notebook, model files, and a final project report.
- Final Presentation: Summarizing project outcomes to stakeholders.

4.2 Out-of-Scope Activities

- Full-scale deployment into FACE Prep's production HRIS.
- Development of specific HR intervention policies (recommendations will be data-driven insights only).
- Collection of new primary data beyond the provided dataset.
- Integration with other internal FACE Prep systems not directly related to this project.

5 Methodology

The project will follow a standard data science lifecycle methodology:

1. **Project Initiation & Planning:** Defining scope, objectives, and deliverables.
2. **Data Acquisition & Understanding:** Loading and performing initial exploration of the dataset.
3. **Data Preprocessing & Cleaning:** Preparing the data for analysis and modeling.
4. **Exploratory Data Analysis (EDA):** Uncovering patterns, anomalies, and relationships in the data.
5. **Feature Engineering & Selection:** Creating and selecting relevant features.
6. **Model Development & Training:** Building and training predictive models.
7. **Model Evaluation & Iteration:** Assessing model performance and refining as necessary.
8. **GUI Development:** Designing and implementing the Streamlit interactive dashboard.
9. **Insight Generation & Reporting:** Documenting findings and preparing the final report and presentation.
10. **Project Closure & Handover:** Delivering all project artifacts.

6 Key Deliverables

- **Comprehensive Jupyter Notebook:** Detailing all data processing, EDA, modeling steps, and code.
- **Trained Machine Learning Model(s):** Saved model file(s) (e.g., using `joblib` or `pickle`).
- **Interactive Streamlit Application:** A functional GUI for predictions and visualization of insights. Source code for the Streamlit app.
- **Final Project Report:** A formal document (this charter serves as the initial version, a final report will detail results) summarizing methodology, findings, model performance, key attrition drivers, and actionable insights.
- **Presentation Slides:** For the final project presentation to HR stakeholders.

7 Success Metrics / Key Performance Indicators (KPIs)

- **Model Performance:** Achieve an F1-score ≥ 0.55 (or Recall ≥ 0.65) for the "Attrition: Yes" class on the unseen test dataset. The chosen metric will be justified based on the business need to correctly identify potential leavers.
- **Insight Generation:** Identification and clear articulation of at least 3-5 statistically significant and actionable drivers of employee attrition.
- **Dashboard Usability:** The Streamlit dashboard should be intuitive and allow HR users to easily obtain predictions and view key insights.
- **Clarity of Communication:** High-quality documentation and clear presentation of findings.
- **Adherence to Project Plan:** Timely completion of deliverables as per the agreed schedule.

8 Tools, Technologies, and Resources

8.1 Software & Libraries

- **Programming Language:** Python 3.x
- **Core Data Science Libraries:** Pandas, NumPy, Scikit-learn
- **Visualization Libraries:** Matplotlib, Seaborn
- **GUI Development Library:** Streamlit
- **Development Environment:** Jupyter Notebook / JupyterLab
- **Version Control (Recommended):** Git / GitHub

8.2 Dataset

- **Primary Dataset:** "IBM HR Analytics Employee Attrition Performance" (Source: Kaggle). Link: https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

9 Project Team Roles

- **Project Sponsor:** Head of Human Resources, FACE Prep Campus.
- **Project Lead / Mentor:** [Instructor's Name / Your Name as Mentor]
- **Data Science Capstone Team Member(s):** [Student's Name(s)] - Responsible for all technical execution, analysis, and development.

10 Timeline Milestones (Illustrative)

- **Week 1:** Project Kick-off, Detailed Planning, Dataset Familiarization.
- **Week 2:** Data Cleaning, Preprocessing, Initial EDA.
- **Week 3:** Advanced EDA, Feature Engineering, Initial Model Building.
- **Week 4:** Model Evaluation, Refinement, Feature Importance Analysis.
- **Week 5:** Streamlit GUI Development, Integration with Model.
- **Week 6:** Finalizing Report, Presentation Preparation, Project Submission.

(A more detailed Gantt chart or project plan may be appended separately)

11 Reporting Communication Plan

- **Weekly Progress Updates:** To Project Lead/Mentor via email or brief meetings.
- **Mid-Project Review (Simulated):** Presentation of EDA findings and preliminary model results.
- **Final Project Presentation:** Formal presentation to simulated HR stakeholders at FACE Prep Campus.

12 Ethical Considerations

- **Data Privacy Anonymity:** Adherence to best practices even with publicly available, anonymized data.
- **Bias Awareness Mitigation:** Conscious effort to identify and discuss potential biases in data and models, and their implications on fairness.
- **Responsible AI Use:** Ensuring that model interpretations and recommendations are presented responsibly, acknowledging model limitations.
- **Transparency:** Clear documentation of the data, methods, and decision-making processes.

13 Learning Outcomes for Capstone Team

- Comprehensive, end-to-end experience in a data science project lifecycle within a simulated corporate environment.
- Practical application of Python data science libraries to solve a real-world business problem.
- Advanced skills in data preprocessing, EDA, feature engineering, predictive modeling, and model evaluation.
- Experience in developing interactive data applications using Streamlit.
- Enhanced ability to interpret model results, derive actionable insights, and communicate technical findings effectively to diverse audiences.

Approvals

[Student's Name(s)] Data Science Capstone Team FACE Prep Campus	[Instructor's Name / Mentor Name] Project Lead / Mentor FACE Prep Campus
Date: _____	Date: _____