HALDIA INSTITUTE OF TECHNOLOGY

A PROJECT REPORT

for

Inhouse Training 2026 (Cluster 3 , batch 7)

on

Employee Attrition Prediction System

Submitted by: **Team INTELLA**

Team members:

1. **JARIN KHAN** (University Roll: 10300223070)
2. **ISHIKA ISHAN** (University Roll: 10300223068)
3. **SADAF AHMAD** (University Roll: 10300223129)

In partial fulfilment of requirements for the award of the degree of

Bachelor of Technology in Information Technology

Under the Guidance of

Mr. Debasish Sahoo

• Abstract :-

Employee attrition is a critical issue faced by organizations worldwide. When employees leave a company, it results in increased recruitment costs, productivity loss, training expenses, and operational disruption. Early prediction of employee attrition allows organizations to take preventive measures and improve employee retention strategies.

This project presents a machine learning–based Employee Attrition Prediction System that predicts whether an employee is likely to leave the organization. The system also identifies key factors contributing to attrition and displays the top reasons behind the prediction.

The system is implemented using:

- Machine Learning model (trained on Employee dataset)
- FastAPI backend for model serving
- React frontend for visualization and prediction interface
- Two trained model files:
  - attrition_model.pkl
  - columns.pkl

The system not only predicts attrition but also provides interpretable insights such as salary, job role, overtime, and satisfaction factors contributing to attrition.

This project demonstrates an end-to-end ML pipeline from data preprocessing and model training to deployment and user-friendly prediction interface.

- # Table of Contents :-

1. Introduction :

1.1 Project Domain :

The project belongs to the domain of Human Resource Analytics and Machine Learning. HR analytics uses data-driven methods to understand employee behavior and improve organizational decisions.

1.2 Motivation :

Employee attrition affects companies through:

- Loss of skilled employees
- Increased hiring and training cost
- Reduced productivity
- Disruption in team performance

Organizations often fail to detect early warning signs of employee resignation. This project aims to build a predictive system that helps HR teams identify employees who may leave the company.

1.3 Objectives :

- To build a machine learning model that predicts employee attrition
- To identify key factors influencing attrition
- To develop a React-based user interface
- To build a FastAPI backend for prediction
- To integrate frontend and backend seamlessly
- To display top reasons for attrition along with prediction

2. Problem Statement :

Organizations struggle to identify employees who are likely to leave. Traditional HR analysis relies on manual observation and historical trends, which is slow and inaccurate.

Problems:

- High employee turnover
- No early warning system
- No automated decision support
- Difficulty identifying key attrition factors

Solution:
Develop a machine learning system that predicts attrition and shows reasons for prediction.

3. Scope of the Project :

The system can be used in:

- HR departments
- IT companies
- Corporate organizations
- Workforce analytics

Future integration possible with:

- Company databases
- HR dashboards
- Employee feedback systems

4. Literature Review :

Research in HR analytics shows that machine learning models can effectively predict employee turnover. Common techniques include:

- Logistic Regression
- Decision Trees
- Random Forest
- Gradient Boosting

Studies indicate that factors like job satisfaction, salary, work-life balance, overtime, and years at company significantly influence attrition.

Traditional HR methods rely on manual observation, whereas machine learning enables automated prediction using historical employee data

5. Dataset Description :

5.1 Dataset Overview :

The dataset contains employee records with features that influence attrition. Each row represents an employee.

5.2 Features Used :

The dataset contains approximately 35 features related to employee demographics, job role, salary, performance, and work environment. These features were used during model training to predict employee attrition.

Feature Categories

Category	Features
Demographic	Age, Gender, Marital Status, Education
Job Information	Job Role, Department, Job Level, Years at Company
Compensation	Monthly Income, Salary Hike Percent
Work Environment	Work Life Balance, Job Satisfaction, Environment Satisfaction
Performance	Performance Rating, Training Times Last Year
Work Behavior	Overtime, Distance From Home
Target	Attrition

5.3 Key UI Features :

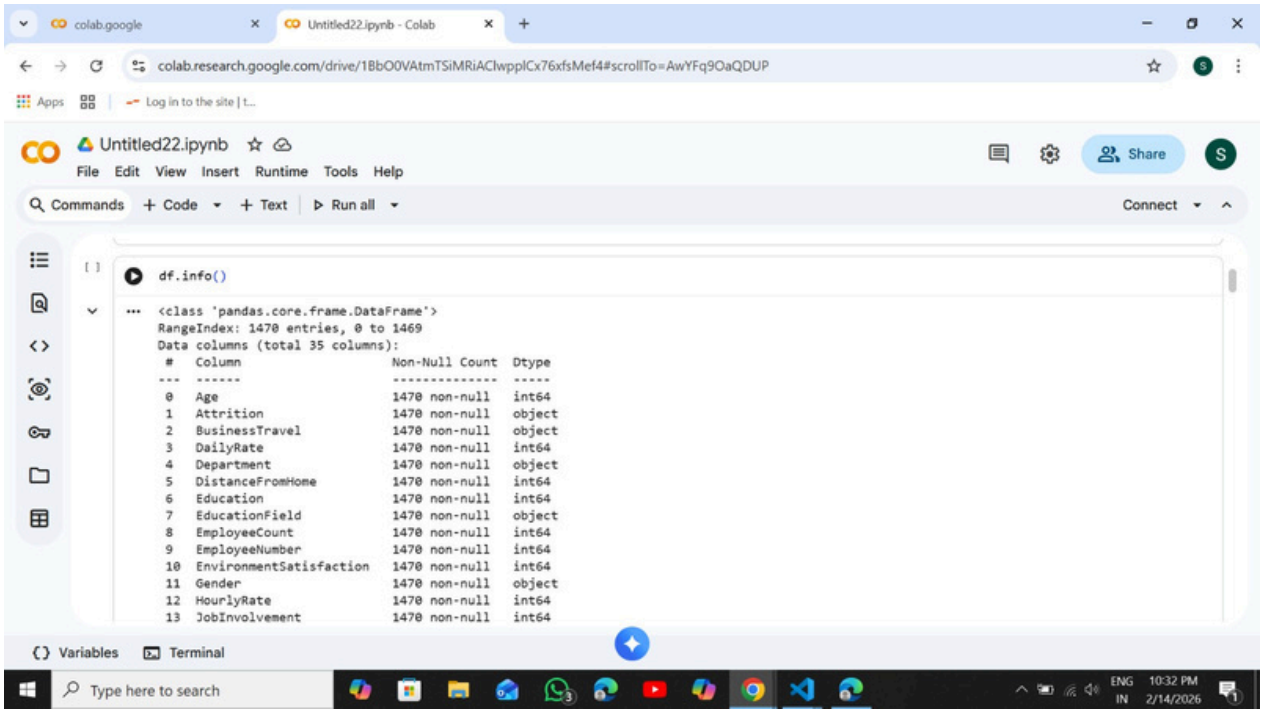
Although the dataset contains many features, only the most influential attributes were selected for the prediction interface to keep the application simple and user-friendly.

The following features are taken as input in the web application

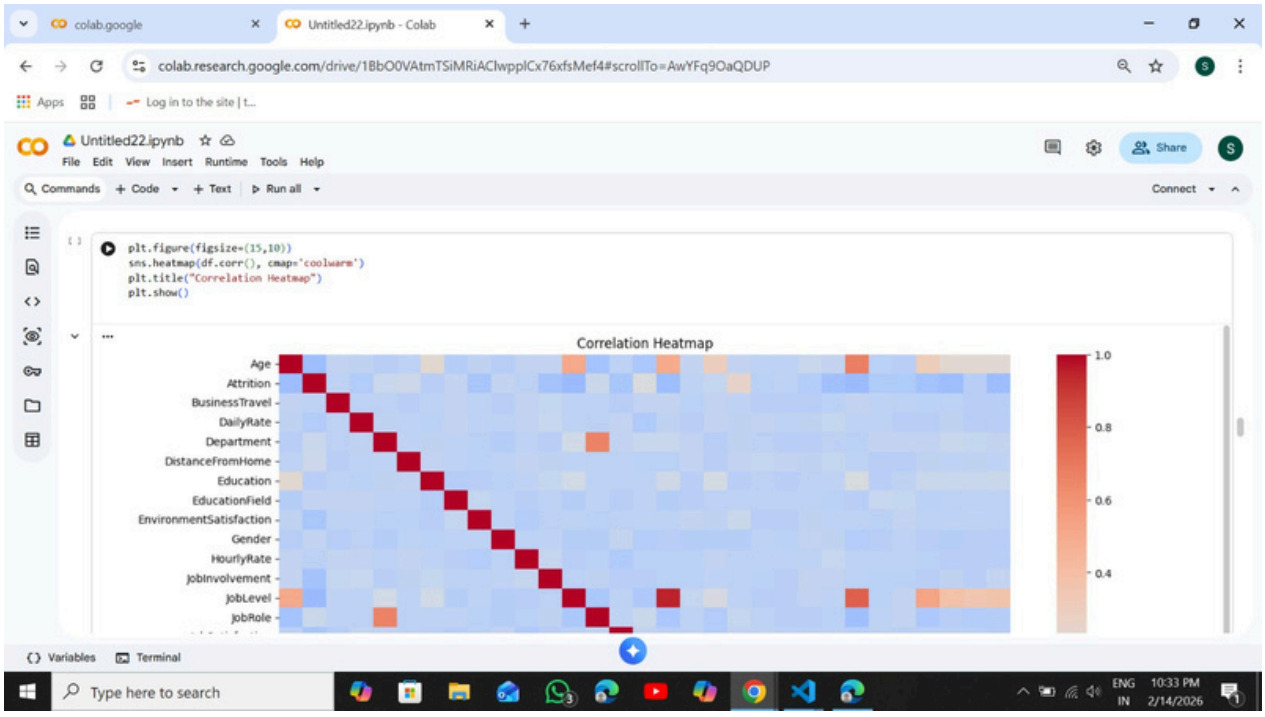
Input Field	Reason for Selection
Age	Basic demographic factor
Monthly Income	Strong impact on attrition
Job Satisfaction	Major attrition indicator
Years at Company	Experience factor
Overtime	High correlation with attrition

These inputs are sufficient for generating predictions because the trained model uses feature importance learned during training.

Data set Information



Correlation Heatmap of data



6. System Architecture and Design :

6.1 System Architecture :

The proposed Employee Attrition Prediction System follows a client–server architecture where the user interface runs on the client side and the prediction logic operates on the server side.

The frontend is developed using React.js, which provides a responsive and interactive interface for HR users to enter employee information and obtain predictions. The frontend communicates with the backend using REST API calls.

The backend is implemented using FastAPI, a high-performance Python web framework suitable for machine learning applications. When the user submits employee details, the frontend sends a request to the backend API. The backend loads the trained machine learning model and processes the input data to generate predictions.

At the core of the system lies the trained machine learning model, which predicts whether an employee is likely to leave the organization. Along with prediction, the model also identifies the most influential features contributing to attrition.

The result is returned to the frontend and displayed to the user in a clear and understandable format.

6.2 High-Level Module Description:

The system consists of the following major modules:

User Interface Module

Provides an interactive web interface where users can enter employee details such as age, salary, job satisfaction, years at company, and overtime status.

Prediction Module

Handles communication between frontend and backend. It sends user input to the backend API and receives prediction results.

Machine Learning Module
Loads the trained model (attrition_model.pkl) and processes the input data to generate predictions and feature importance scores.

API Module
Developed using FastAPI, this module receives requests from the frontend and returns predictions in JSON format.

Explanation Module
Identifies the top contributing factors responsible for attrition risk and sends them to the frontend for display.

7. Methodology and Implementation:

7.1 Data Processing :

The dataset used in this project contains multiple employee-related attributes. Before training the model, the following preprocessing steps were performed:

- Handling missing values
- Encoding categorical variables
- Feature selection
- Data normalization
- Train-test split

These steps ensured that the dataset was clean and suitable for machine learning training.

7.2 Model Training :

Several classification algorithms were tested during development, including Logistic Regression and Random Forest.

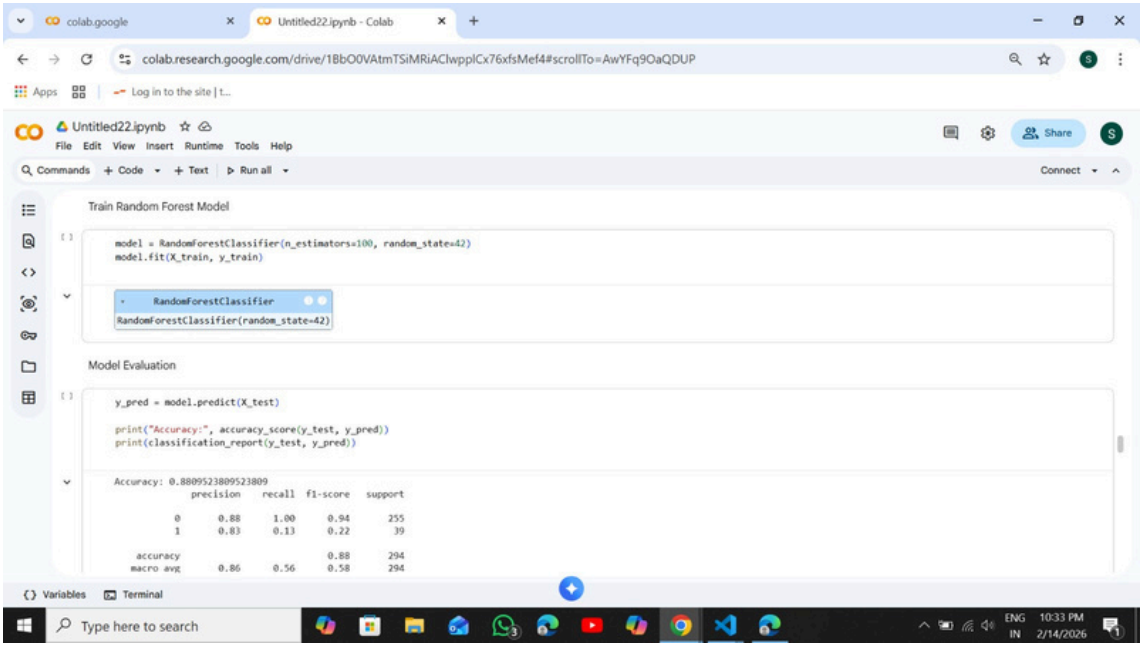
The Random Forest Classifier provided better performance and was selected for final implementation due to its:

- Higher accuracy
- Ability to handle nonlinear relationships
- Feature importance extraction

The trained model was saved using the Joblib library as:

- attrition_model.pkl
- columns.pkl

These files are loaded during runtime in the backend.



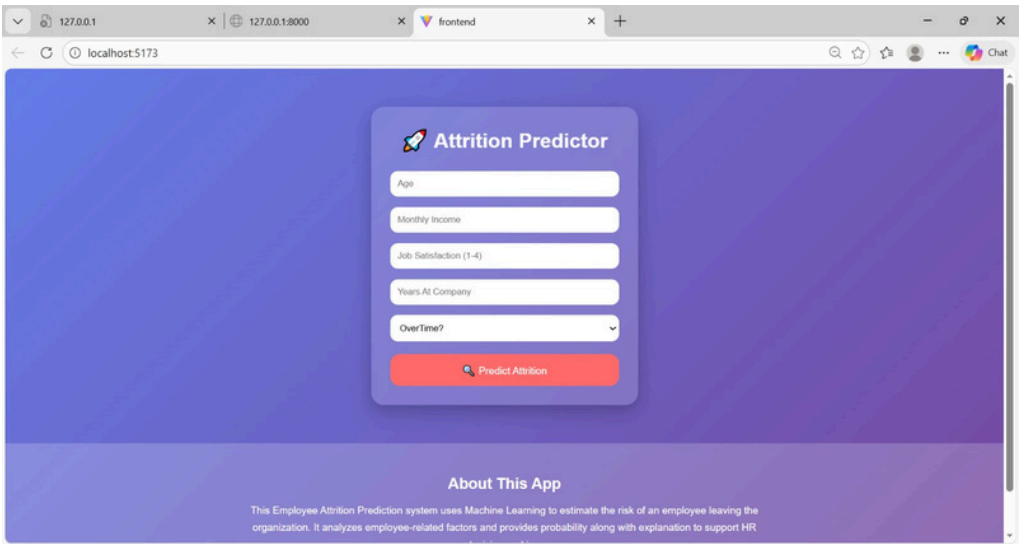
7.3 User Interface and Configuration :

The system provides a simple and modern interface designed for HR users. The UI allows users to enter employee details and obtain predictions instantly.

The interface includes:

- Input form
- Dropdown for overtime
- Predict button
- Result display section
- About section

The design ensures that the application remains easy to use even for non-technical users.



7.4 Prediction Process :

1. User enters employee data in the React interface
2. Data is sent to FastAPI backend
3. Backend loads trained model
4. Model predicts attrition risk
5. Feature importance is calculated
6. Top reasons identified
7. Result returned to frontend
8. UI displays prediction and reasons

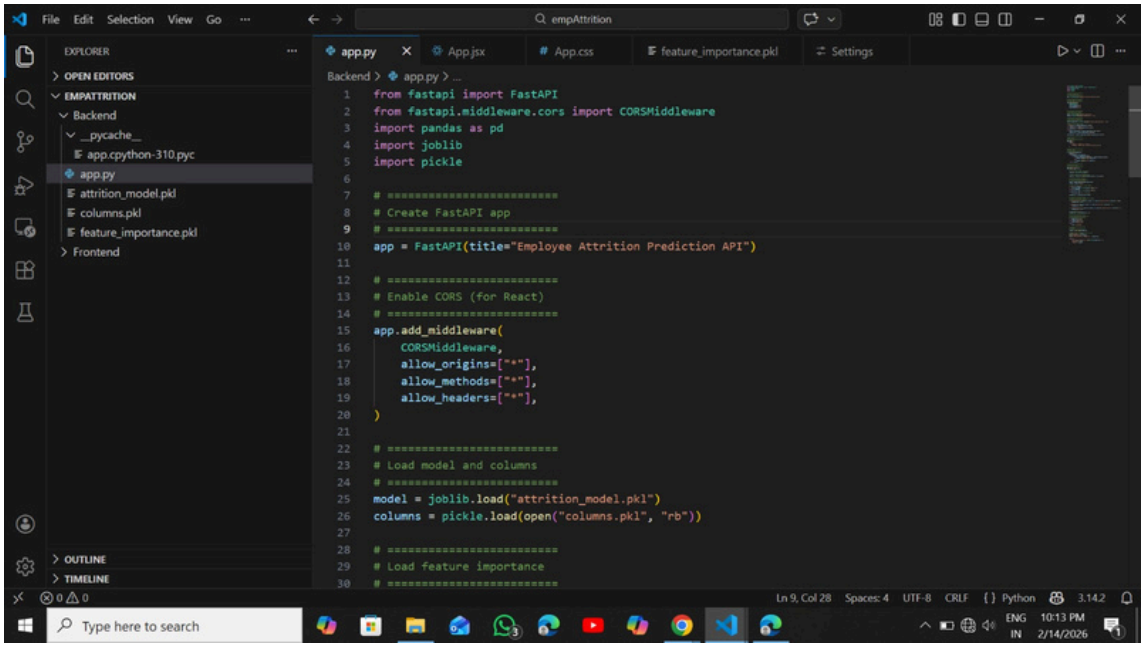
7.5 Key Features of the System :

1. Predicts employee attrition risk
2. Displays key reasons for attrition
3. User-friendly interface
4. Fast prediction using ML model
5. Real-time API communication
6. Scalable architecture

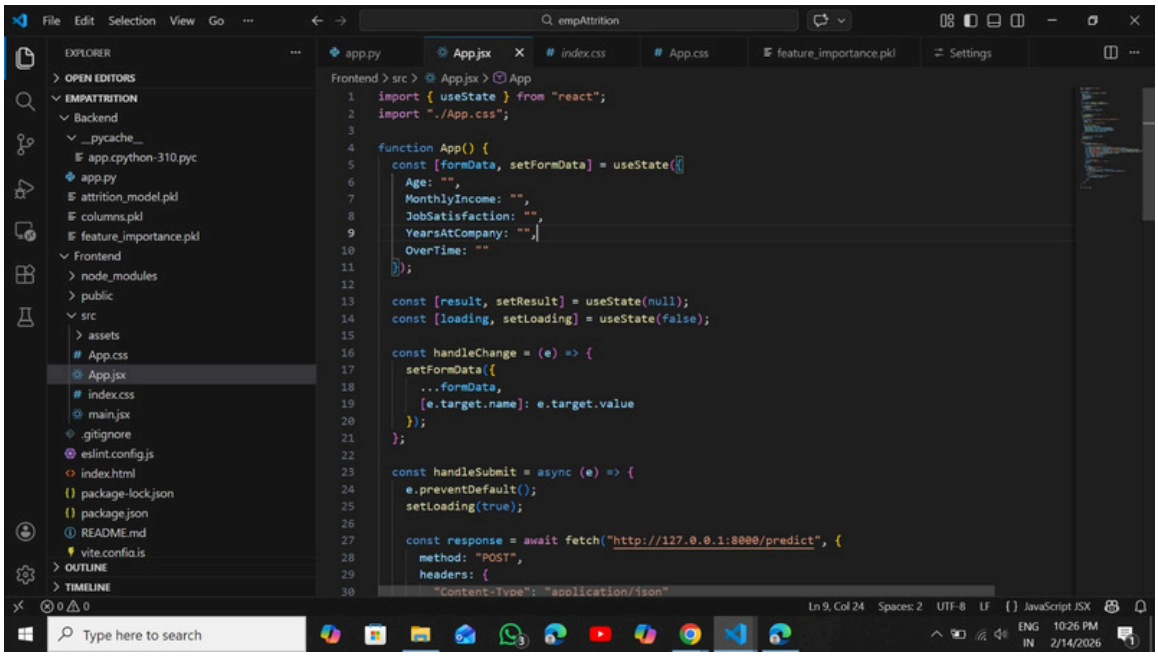
7.6 Technologies Used :

Component	Technology
Frontend	React.js
Backend	FastAPI
Machine Learning	Scikit-learn
Programming Language	Python
Model Storage	Joblib
API Testing	FastAPI Docs

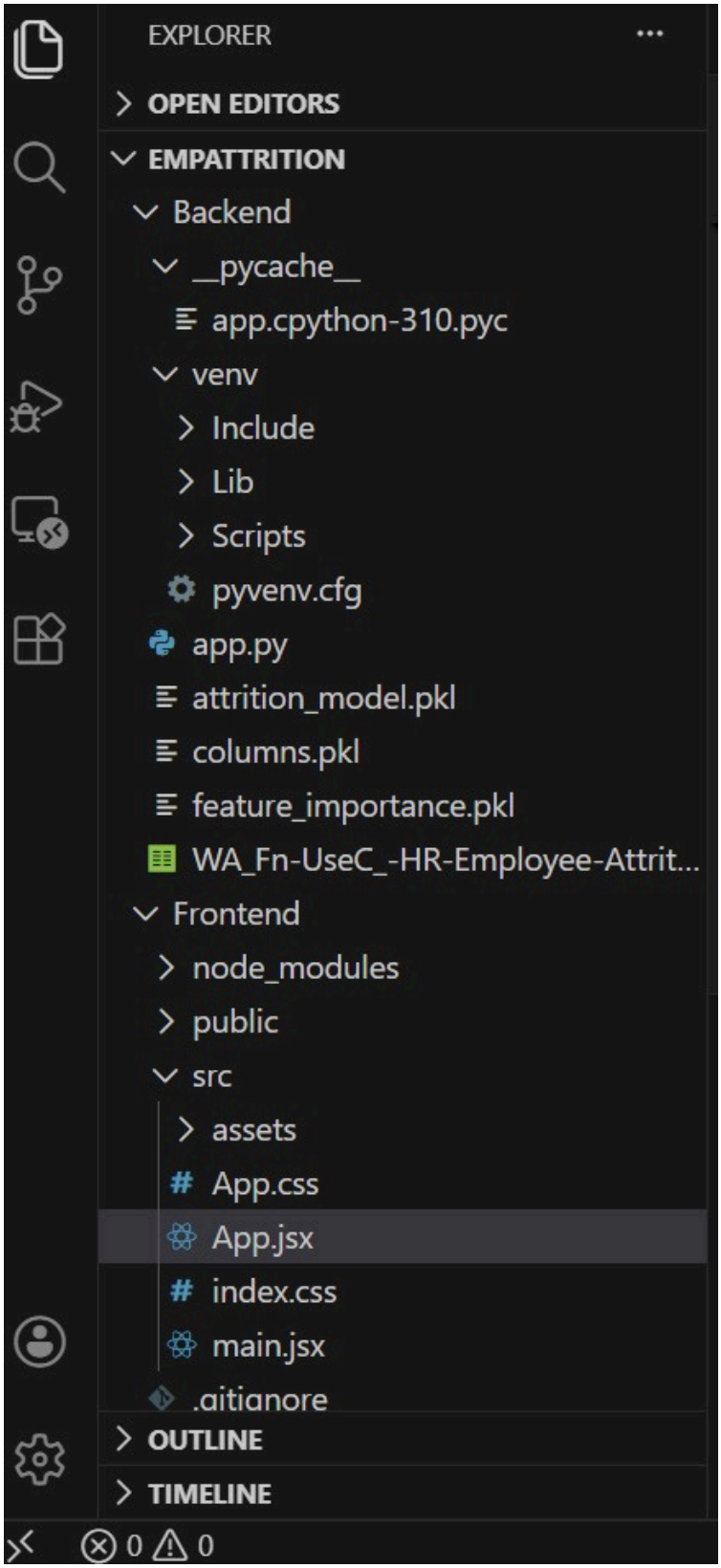
- **Backend (app.py)**



- **Frontend (app.jsx)**

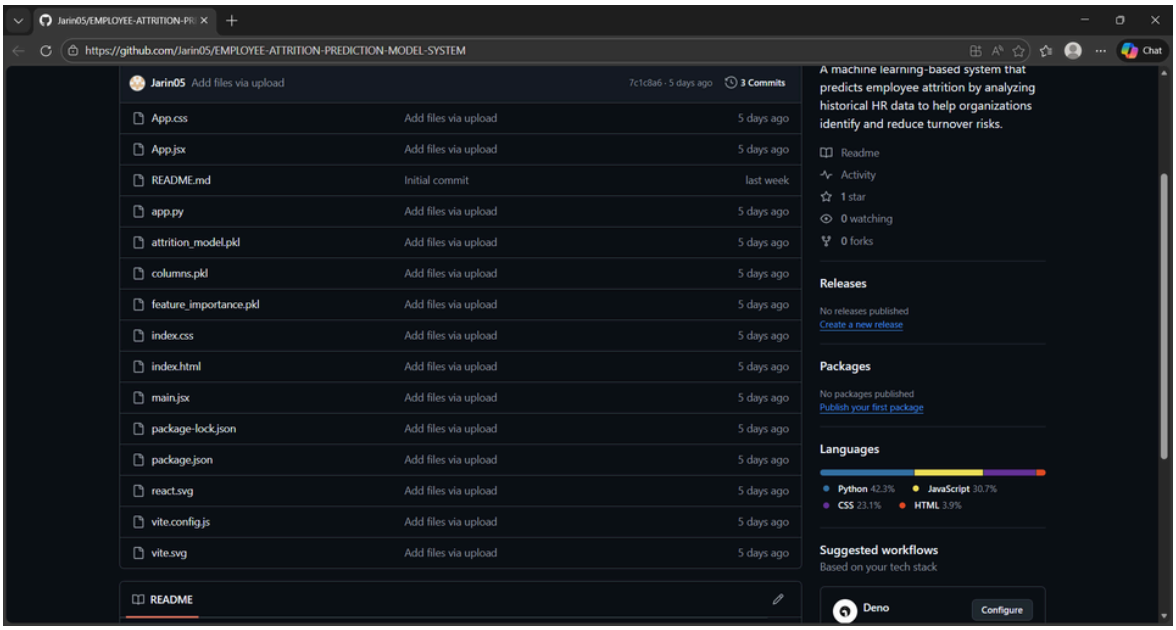


8. Project Folder Structure ：



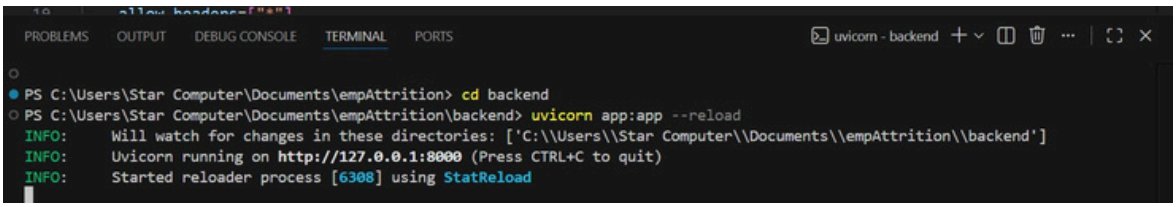
9. Deployment:

10.1 GitHub :

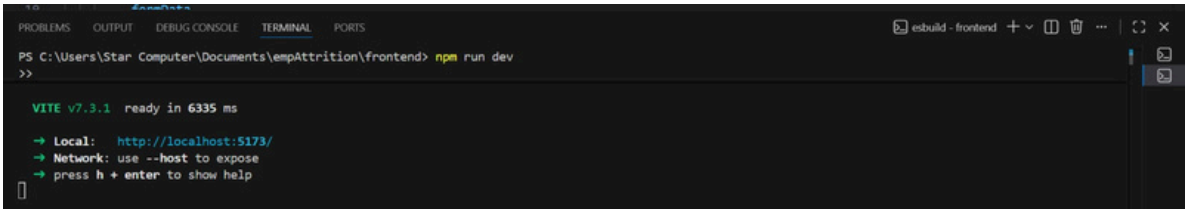


10.2 Local Deployment :

- **Backend :** uvicorn main :app --reload



- **Frontend :** npm start



10. Results and Discussion:

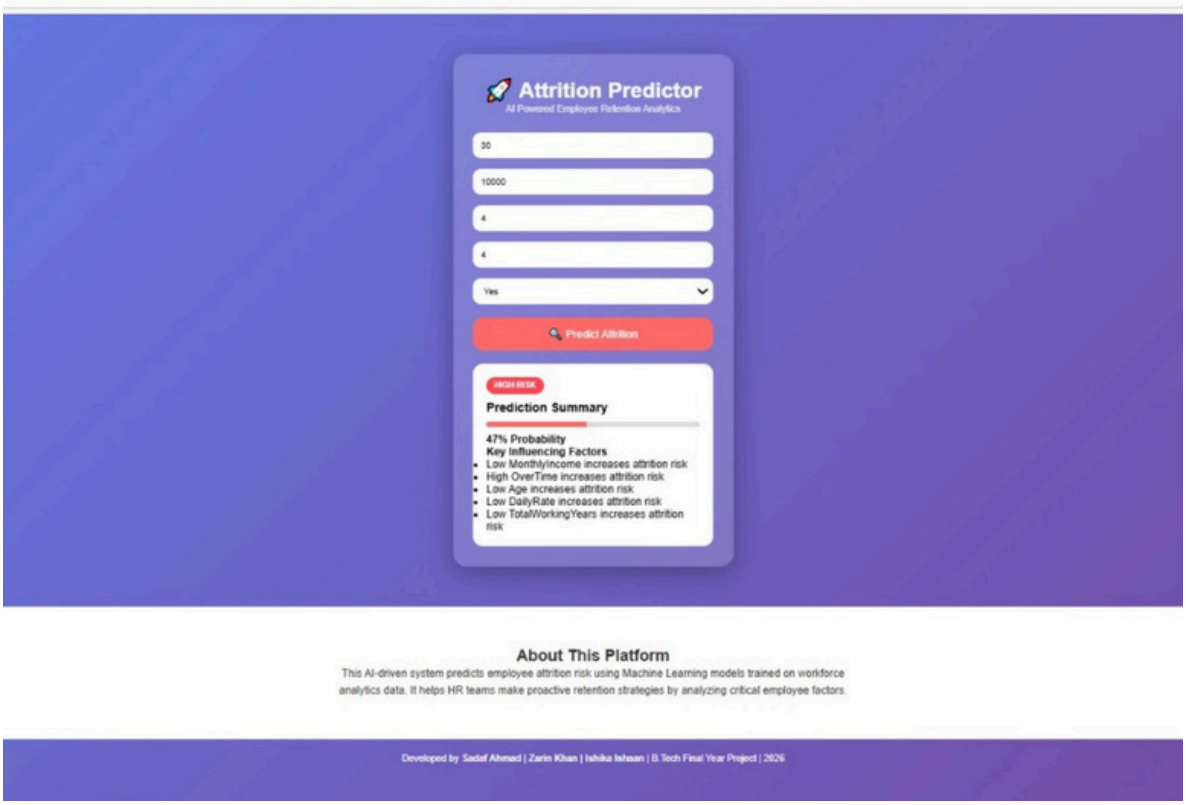
The system was tested using multiple input combinations to evaluate prediction accuracy and reliability. The model successfully predicted employee attrition risk based on input parameters.

The application was able to:

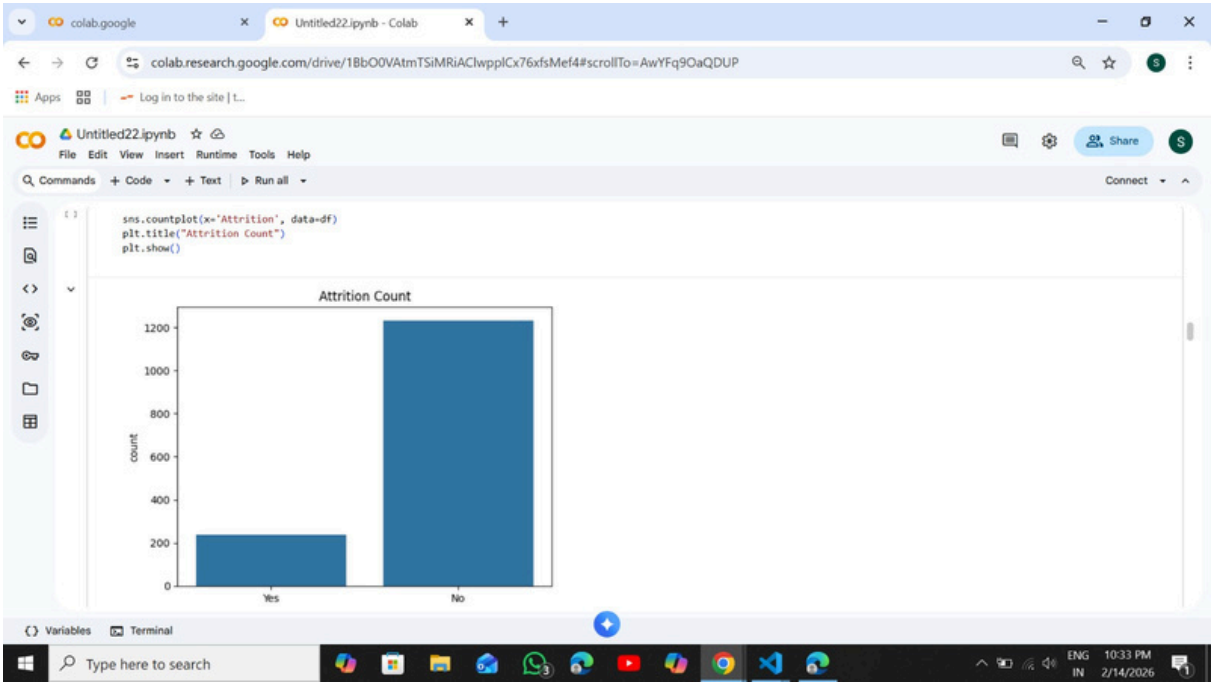
- Generate predictions in real time
- Display meaningful reasons
- Provide consistent results

Testing showed that employees with high overtime, low job satisfaction, and low salary were more likely to leave, which aligns with real-world HR trends.

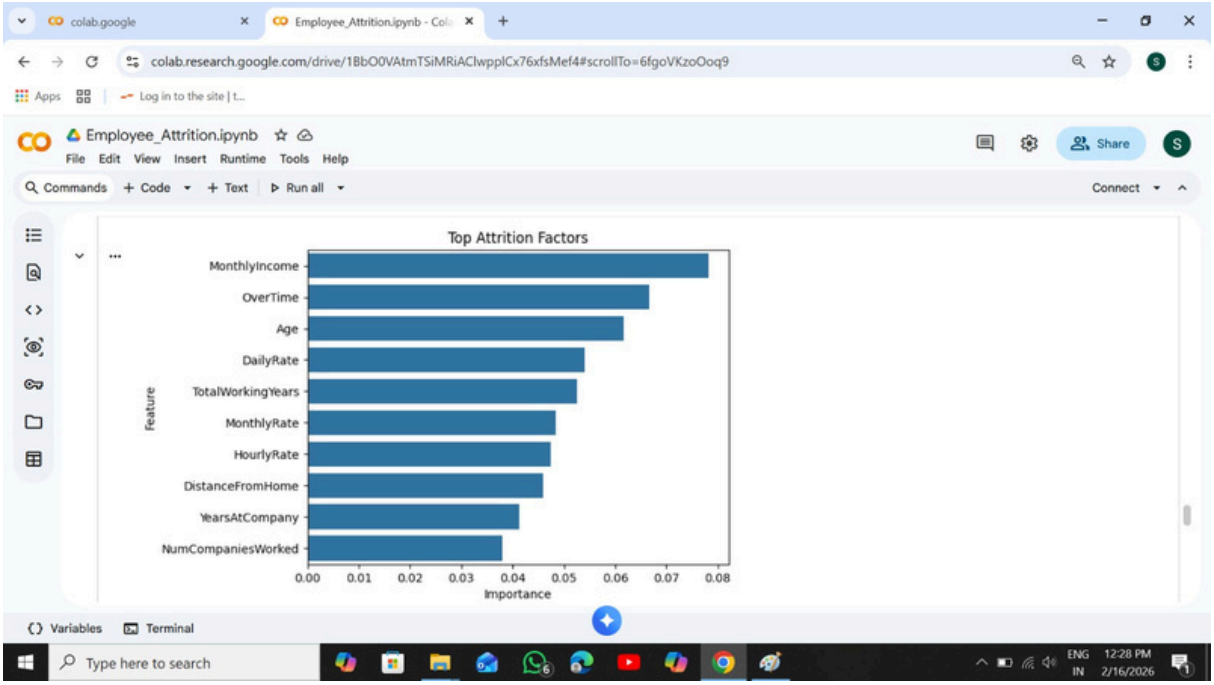
The user interface also performed smoothly, allowing users to enter data and receive predictions without delay.



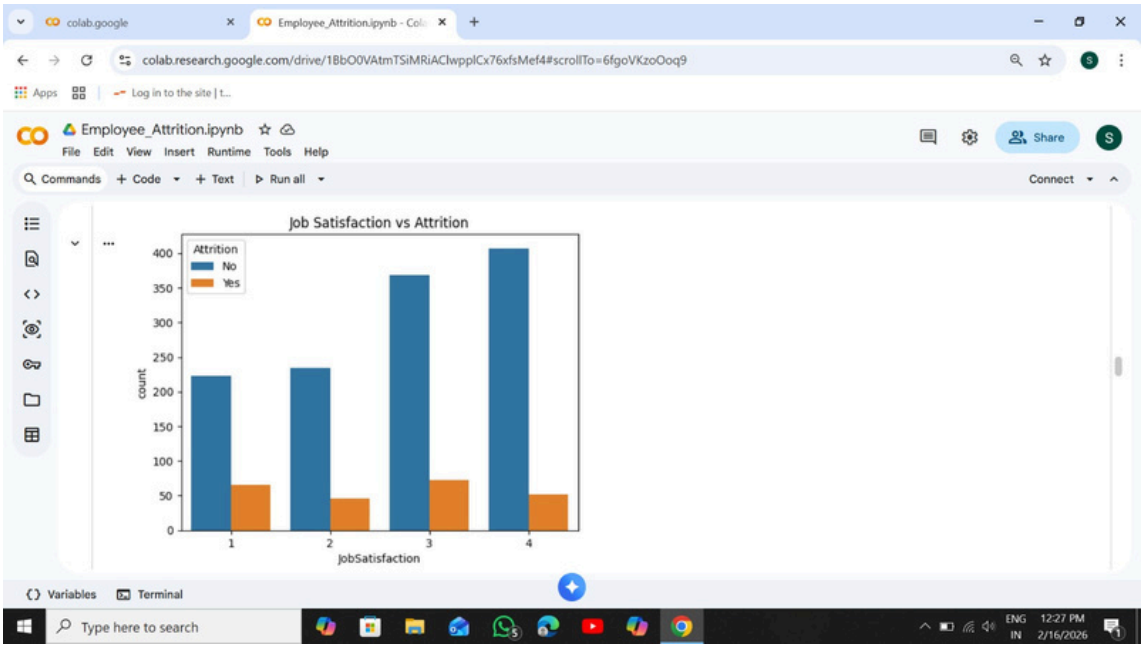
Attrition Count from data set



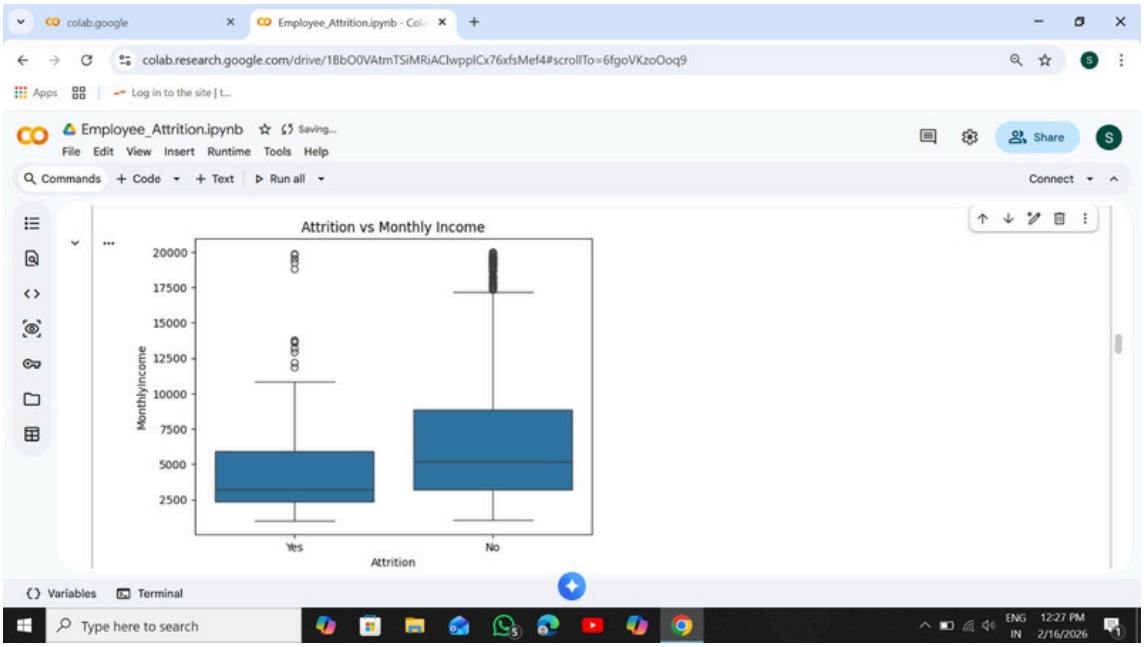
Top attrition factors



Job Satisfaction vs Attrition



Attrition vs Monthly Income



11. Advantages of the System :

- **Early Attrition Detection:** Helps organizations identify employees who are likely to leave in advance.
- **Data-Driven Decision Making:** Provides HR teams with insights based on employee data rather than assumptions.
- **User-Friendly Interface:** Simple React-based interface allows non-technical users to operate the system easily.
- **Real-Time Prediction:** Generates instant predictions after entering employee details.
- **Reason Explanation:** Displays key factors responsible for attrition, improving transparency and understanding.
- **Scalable Architecture:** The React + FastAPI structure can be extended for larger datasets and real company use.
- **Automation of Analysis:** Reduces manual HR effort in analyzing employee behavior.
- **Educational Value:** Demonstrates practical implementation of machine learning in HR analytics.
- **Modular Design:** Frontend, backend, and ML model are separated, making future updates easier.
- **Lightweight Deployment:** Can run locally without heavy infrastructure requirements.

12. Limitations of the System :

- **Limited Dataset:** The model is trained on a sample dataset and may not fully represent real company scenarios.
- **Input Dependency:** Predictions depend on the accuracy of user-entered data.
- **Not Connected to Real HR Database:** The system currently uses manual input rather than live employee records.
- **Limited Feature Inputs in UI:** Only a subset of features is used in the interface for simplicity.
- **Model Generalization:** Performance may vary when applied to different organizations or industries.
- **No Continuous Learning:** The model does not update automatically with new employee data.
- **Basic Visualization:** The current system provides textual output rather than detailed analytics dashboards.
- **Local Deployment Only:** The project is currently designed for local use and not deployed on cloud infrastructure.
- **Security Considerations:** Authentication and user management are not implemented in this version.
- **Not a Replacement for HR Judgment:** The system supports decision-making but cannot fully replace human evaluation.

13. Conclusion :

The Employee Attrition Prediction System demonstrates how machine learning can be applied to human resource analytics to predict employee turnover.

The system successfully integrates a React frontend with a FastAPI backend and a trained machine learning model. It not only predicts attrition but also provides explanations for the prediction, making it useful for HR decision-making.

This project provided practical experience in:

- Machine learning model development
- React frontend design
- FastAPI backend development
- API integration
- Full-stack project deployment

The system can be further improved with larger datasets, cloud deployment, and dashboard features.

14. References :

- Scikit-learn Documentation
- FastAPI Documentation
- React Documentation
- Kaggle Employee Attrition Dataset