Microsoft Elevate

CAPSTONE PROJECT

ATTRISENSEAI - DATA DRIVEN INSIGHTS FOR EMPLOYEE RETENTION

PRESENTED BY

STUDENT NAME: Sruthilayaa K

COLLEGE NAME: Adhiyamaan College Of Engineering

DEPARTMENT: Electronics And Communication Engineering

EMAIL ID:sruthilayaa2621@gmail.com

Microsoft Elevate

OUTLINE:

- **PROBLEM STATEMENT**
- **PROPOSED SYSTEM/SOLUTION**
- **SYSTEM DEVELOPMENT APPROACH**
- **ALGORITHM & DEPLOYMENT**
- **RESULT**
- **CONCLUSION**
- **FUTURE SCOPE**
- **REFERENCES**

Microsoft Elevate

PROBLEM STATEMENT:

Employee attrition is one of the most critical challenges faced by organizations across industries. High attrition rates lead to increased recruitment and training costs, loss of organizational knowledge, reduced productivity, and negative impacts on team morale. Traditional methods of addressing employee turnover are often **reactive**, relying on exit interviews and manual analysis, which fail to identify early warning signs of attrition.

With the availability of large volumes of workforce data, there is a growing need for **data-driven and predictive systems** that can proactively identify employees who are at risk of leaving an organization. However, many existing solutions lack interpretability, actionable insights, and integration with modern AI-based decision support systems

PROPOSED SOLUTION:

The proposed system aims to address the challenge of **predicting employee attrition risk** in organizations in order to support proactive retention strategies. By leveraging **data analytics, machine learning, and generative AI**, the system forecasts the likelihood of an employee leaving the organization and provides actionable recommendations to reduce attrition. The solution consists of the following components:

1. Data Collection

- Gather historical employee data including demographic details, job roles, compensation, work experience, satisfaction levels, and performance-related attributes.
- Use structured datasets containing employee records with labeled attrition outcomes to enable supervised learning.
- Ensure data consistency by aligning features with industry-standard HR datasets (e.g., IBM HR Analytics dataset).

2. Data Preprocessing

- Clean and preprocess the collected data to handle missing values, duplicate entries, and inconsistencies.
- Encode categorical variables such as department, job role, education field, and business travel using appropriate encoding techniques.
- Normalize and scale numerical features such as income, years of experience, and salary hike percentage.
- Perform feature selection and engineering to identify key factors influencing employee attrition.

PROPOSED SOLUTION:

2. Data Preprocessing

- Clean and preprocess the collected data to handle missing values, duplicate entries, and inconsistencies.
- Encode categorical variables such as department, job role, education field, and business travel using appropriate encoding techniques.
- Normalize and scale numerical features such as income, years of experience, and salary hike percentage.
- Perform feature selection and engineering to identify key factors influencing employee attrition.

3. Machine Learning Algorithm

- Implement a **Logistic Regression–based machine learning pipeline** to predict the probability of employee attrition.
- Use preprocessing pipelines (e.g., ColumnTransformer) to ensure consistent feature handling during training and inference.
- The model outputs both:
  - A **binary classification** (High Risk / Low Risk of Attrition)
  - A **probability score** representing attrition likelihood
- The simplicity and interpretability of Logistic Regression make it suitable for HR decision-making and explainability.

PROPOSED SOLUTION:

4. AI-Driven Recommendation System

- Integrate **Generative AI (Google Gemini)** to provide intelligent, ethical, and context-aware retention recommendations.
- Use the predicted attrition probability and key employee indicators as input prompts for the AI model.
- Generate actionable suggestions such as improving work-life balance, career growth opportunities, compensation adjustments, and employee engagement initiatives.

5. Deployment

- Develop a **user-friendly web application** using Streamlit that allows HR professionals to input employee details and view predictions in real time.
- Visualize attrition risk using an intuitive **gauge (speedometer)** for better interpretability.
- Deploy the application on a **cloud-based platform (Microsoft Azure)** to ensure scalability, availability, and ease of access.
- Secure sensitive credentials (e.g., API keys) using environment variables and cloud configuration settings.

PROPOSED SOLUTION:

6. Evaluation

- Evaluate the machine learning model using performance metrics such as:
  - Accuracy
  - Precision
  - Recall
  - F1-score
- Analyze classification reports and confusion matrices to assess prediction reliability.
- Continuously monitor model performance and retrain the system as new employee data becomes available.

SYSTEM APPROACH:

The System Approach outlines the overall strategy and methodology adopted for designing, developing, and implementing the **Employee Attrition Prediction and Retention Advisory System**. The system integrates data preprocessing, machine learning–based prediction, and AI-driven decision support within a scalable web application framework.

1. System Requirements

- The proposed system requires both **hardware** and **software** components to ensure smooth execution, scalability, and reliable performance.
- **Hardware Requirements**
- Processor: Intel i5 or higher (or equivalent)
- RAM: Minimum 8 GB
- Storage: Minimum 10 GB free disk space
- Internet connectivity for cloud deployment and AI services.

SYSTEM APPROACH:

- **Software Requirements**
- Operating System: Windows / Linux / macOS
- Programming Language: Python 3.10 or above
- Development Environment: VS Code / Jupyter Notebook
- Web Framework: Streamlit
- Cloud Platform: Microsoft Azure (App Service)
- Version Control: Git & GitHub

2. Libraries Required to Build the Model

- The following Python libraries are used for data processing, model development, deployment, and AI integration:
- **Data Handling & Analysis**
- pandas – for data manipulation and analysis
- numpy – for numerical computations

Microsoft Elevate

SYSTEM APPROACH:

- **Data Preprocessing & Machine Learning**
- scikit-learn – for:
  - Logistic Regression model
  - Feature scaling
  - Encoding categorical variables
  - Pipeline and ColumnTransformer
  - Model evaluation metrics
- **Model Persistence**
- joblib – for saving and loading the trained machine learning pipeline
- **Web Application Development**
- streamlit – for building an interactive and user-friendly web interface
- **Data Visualization**
- plotly – for creating interactive visualizations such as the attrition risk gauge
- **AI Integration**
- google-genai – for integrating Google Gemini to generate AI-driven retention suggestions
- **Environment & Configuration**
- python-dotenv – for securely managing environment variables (API keys)

ALGORITHM & DEPLOYMENT:

This section describes the machine learning algorithm used for predicting employee attrition and explains how the trained model is deployed as a web-based application for real-time inference and decision support.

1. Algorithm Selection

- The proposed system uses **Logistic Regression** as the primary machine learning algorithm for predicting employee attrition.
- Logistic Regression is a supervised classification algorithm that estimates the probability of a binary outcome—in this case, whether an employee is likely to leave the organization or not. It was selected based on the following considerations:
- The problem involves **binary classification** (Attrition: Yes / No).
- The dataset consists of both **categorical and numerical features**, which are well supported by Logistic Regression when combined with proper preprocessing.
- Logistic Regression offers **high interpretability**, making it suitable for HR analytics where explainability is critical.
- The algorithm performs efficiently on structured tabular data and provides probabilistic outputs, which are useful for risk assessment.

ALGORITHM & DEPLOYMENT:

2. Data Input

- The algorithm uses a set of employee-related features derived from historical HR data. These inputs capture demographic, professional, compensation, and satisfaction-related attributes that influence attrition.
- Key input features include:
- Demographic attributes: Age, Gender, Marital Status, Education Level
- Job-related attributes: Department, Job Role, Job Level, Total Working Years
- Compensation factors: Monthly Income, Percent Salary Hike, Stock Option Level
- Work conditions: Business Travel, OverTime, Years at Company
- Satisfaction indicators: Job Satisfaction, Environment Satisfaction, Work-Life Balance, Job Involvement
- Career growth indicators: Years Since Last Promotion, Years in Current Role
- These features are preprocessed using encoding and scaling techniques before being passed to the model.

ALGORITHM & DEPLOYMENT:

3. Training Process

- The Logistic Regression model is trained using historical employee data with known attrition outcomes.
- The training process includes:
- Data preprocessing using a **Pipeline and ColumnTransformer** to ensure consistent handling of numerical and categorical features.
- Splitting the dataset into training and testing sets to evaluate generalization performance.
- Model fitting using the training data to learn the relationship between employee attributes and attrition outcomes.
- Performance evaluation using metrics such as accuracy, precision, recall, F1-score, and classification reports.
- The trained pipeline (including preprocessing and model) is saved using joblib for reuse during deployment.

ALGORITHM & DEPLOYMENT:

4. Prediction Process

- During the prediction phase:
- User-provided employee details are collected through a web interface.
- The input data is passed to the trained machine learning pipeline, which automatically applies preprocessing steps.
- The model outputs:
  - A **binary prediction** indicating high or low risk of attrition.
  - A **probability score** representing the likelihood of attrition.
- The probability score is visualized using a gauge chart to enhance interpretability.
- Additionally, the predicted attrition risk and key employee indicators are passed to a **Generative AI model (Google Gemini)**, which generates ethical and practical retention suggestions to help mitigate attrition risk.

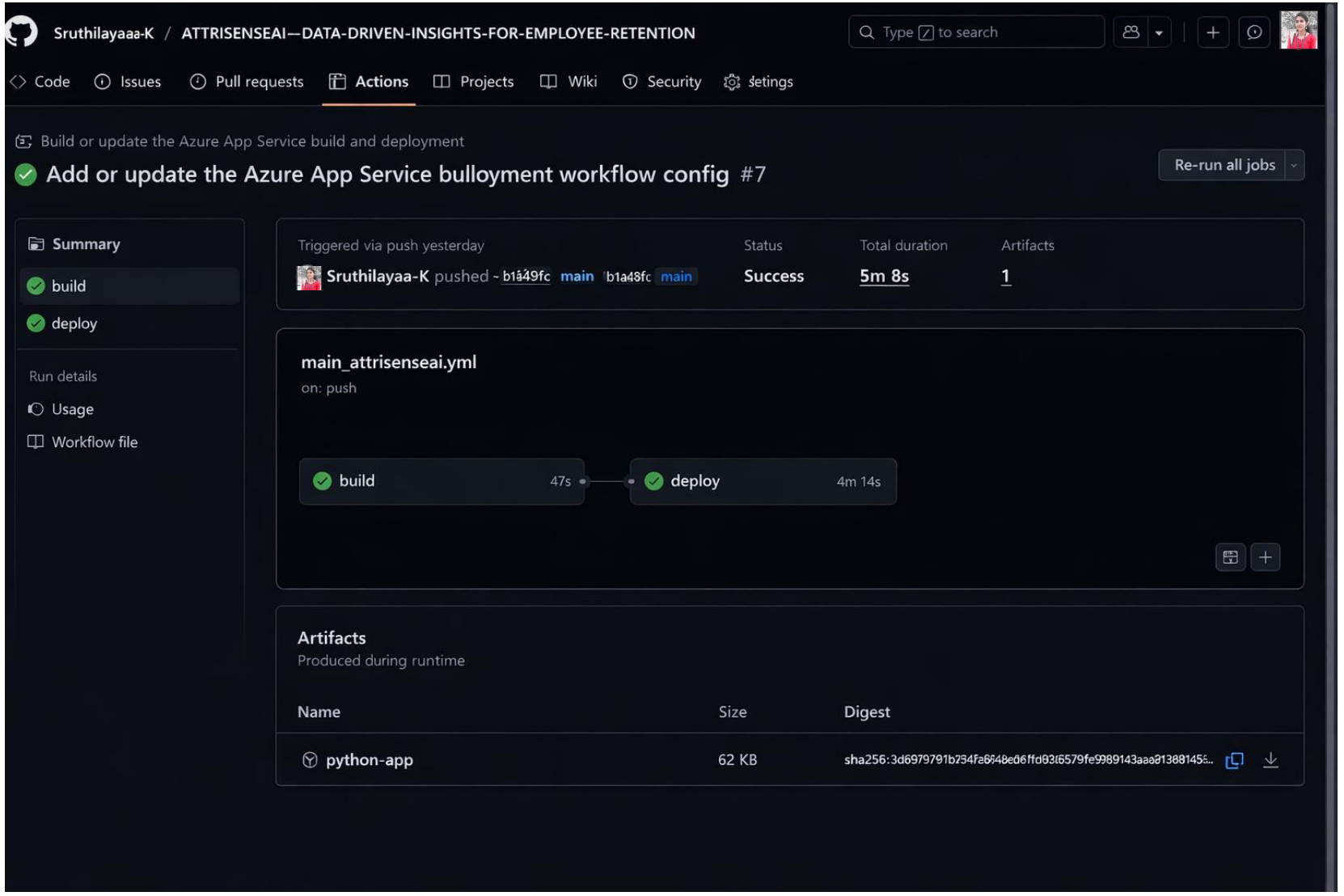
ALGORITHM & DEPLOYMENT:

5. Deployment

- The trained model is deployed as an interactive web application using **Streamlit** and hosted on **Microsoft Azure App Service**.
- Key deployment aspects include:
- A user-friendly interface for real-time data input and prediction.
- Secure handling of sensitive credentials using environment variables.
- Scalable cloud infrastructure to support multiple users.
- Continuous deployment enabled through GitHub integration.
- This deployment approach ensures accessibility, scalability, and real-time decision support for HR professionals.

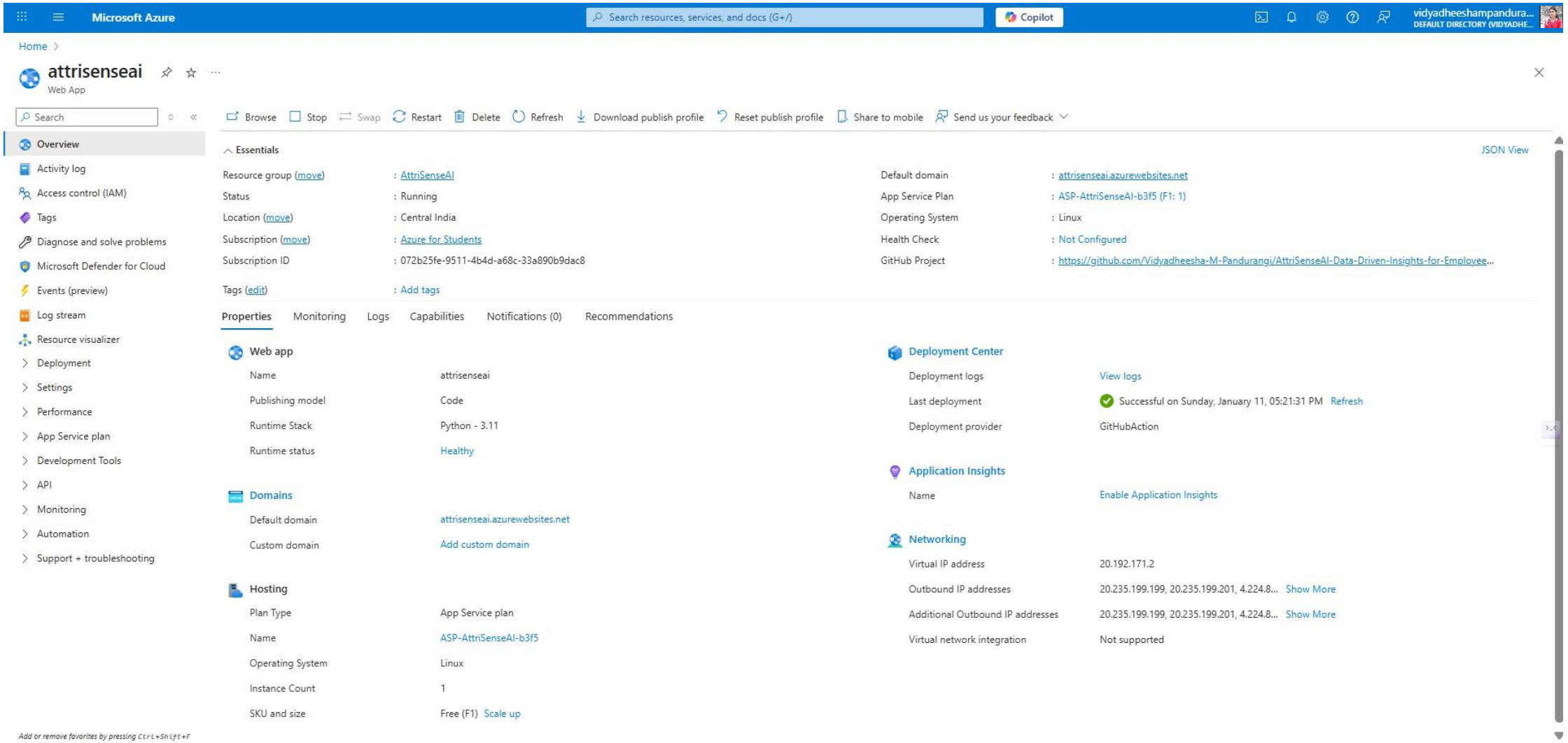
Microsoft Elevate

RESULT:



Microsoft Elevate

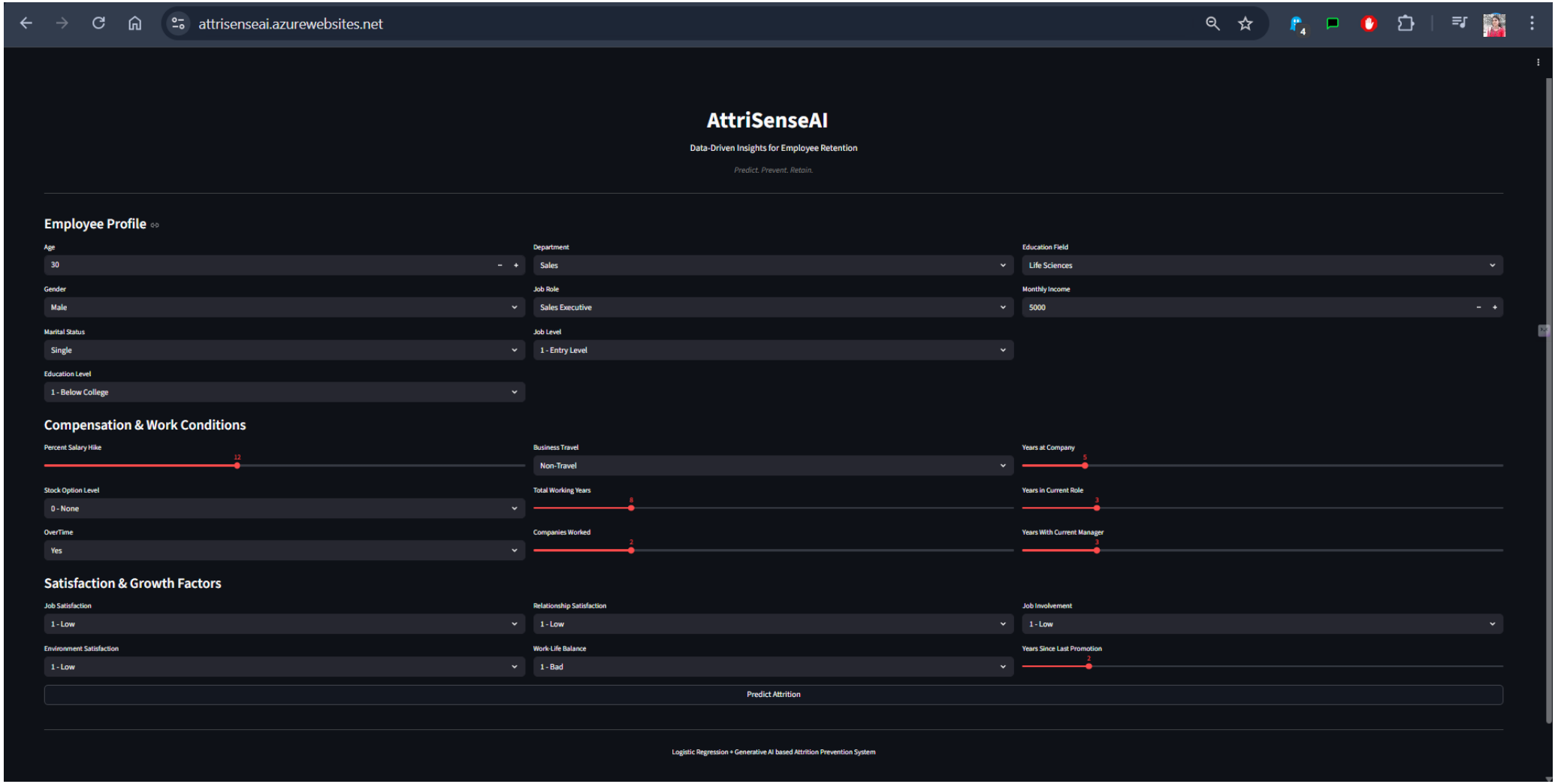
RESULT:



Configurations of Deployed Application – Using Microsoft Azure Cloud

Microsoft Elevate

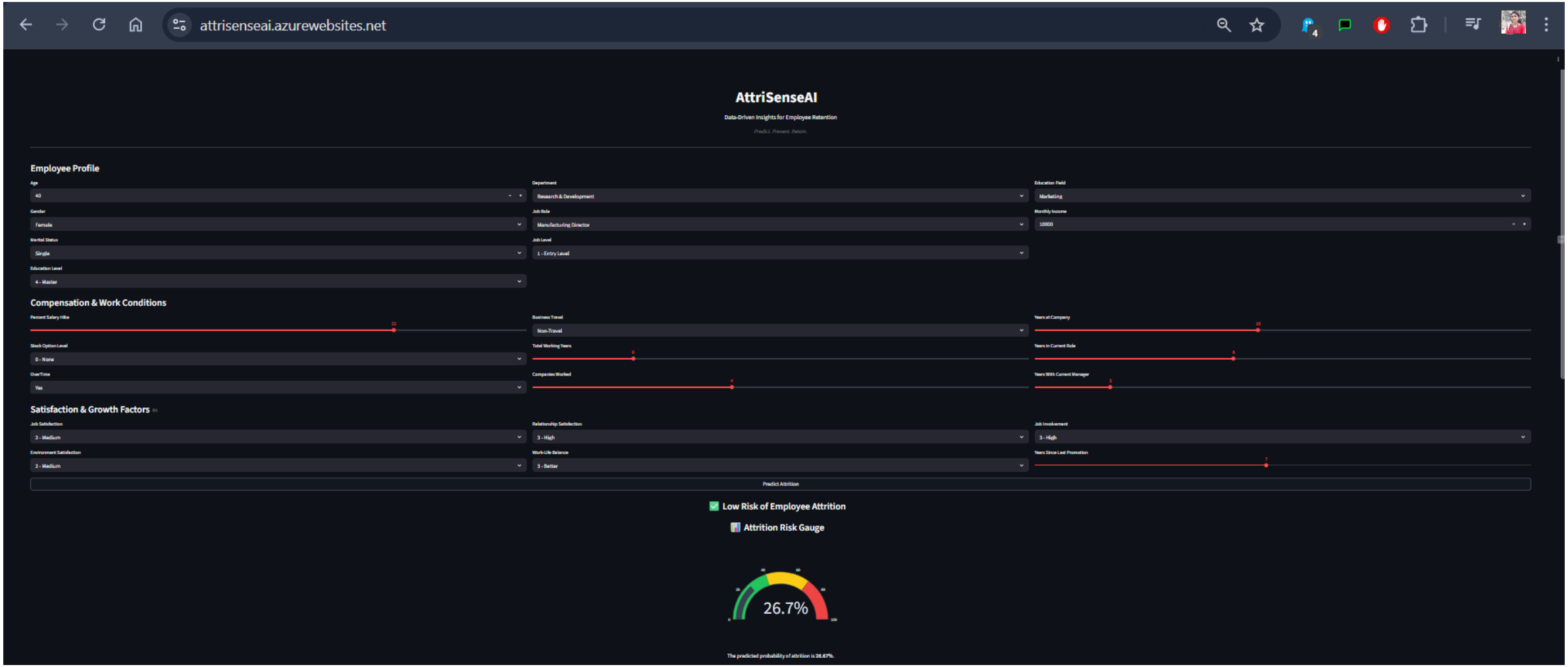
RESULT:



UI of "AttriSenseAI" – Deployed in Microsoft Azure Cloud

Microsoft Elevate

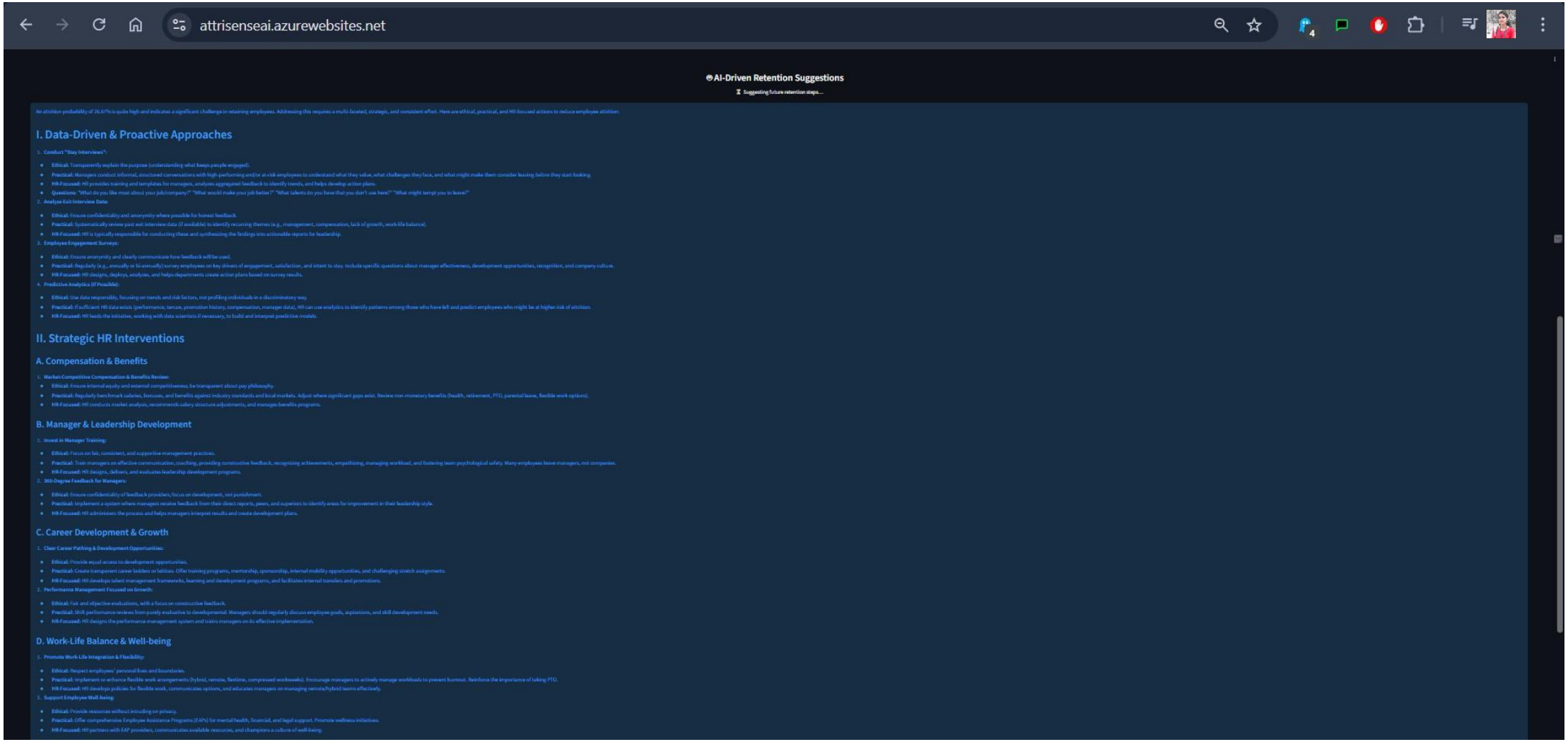
RESULT:



Sample Prediction

Microsoft Elevate

RESULT:



AI Suggestions (By Google Gemini) – Based on the Predictions

Microsoft Elevate

CONCLUSION:

The proposed system successfully demonstrates the effectiveness of using machine learning and cloud-based AI technologies to predict employee attrition and support proactive retention strategies. By leveraging a Logistic Regression model, the system accurately identifies employees at risk of leaving based on demographic, job-related, compensation, and satisfaction factors. The integration of an interactive web interface and AI-driven recommendations enhances usability and provides actionable insights for HR professionals. Deployment on Microsoft Azure ensures scalability, accessibility, and reliability, making the solution suitable for real-world organizational environments.

During implementation, challenges such as data preprocessing consistency, model version compatibility, and AI service availability were encountered and addressed through robust pipelines, secure environment configuration, and exception handling mechanisms. Future improvements may include adopting advanced machine learning models, enhancing explainability, and integrating real-time HR data. Accurate attrition prediction plays a crucial role in maintaining a stable and productive workforce, as it enables organizations to reduce turnover costs, retain valuable talent, and improve overall employee satisfaction through timely and informed interventions.

FUTURE SCOPE:

The proposed employee attrition prediction system can be further enhanced by incorporating additional data sources such as real-time employee engagement surveys, performance evaluation records, and organizational policy changes. Integrating data from multiple departments or business units would improve the model's ability to capture complex workforce dynamics and increase prediction accuracy. The system can also be expanded to support multiple organizations or regions, enabling comparative analysis of attrition trends across different locations and industries.

From a technological perspective, future improvements may include optimizing the machine learning algorithm using advanced techniques such as ensemble models, gradient boosting, or deep learning to improve predictive performance. The integration of emerging technologies like **edge computing** can enable faster, localized inference for large enterprises, while continuous learning pipelines on **Microsoft Azure** can support automated model retraining. Additionally, enhancing explainability, real-time dashboards, and AI-driven policy simulations would further strengthen the system's value as an intelligent decision-support tool for workforce management.

Microsoft Elevate

REFERENCES:

- **Microsoft Corporation:** Microsoft Azure App Service Documentation. Available at: https://learn.microsoft.com/azure/app-service

- **Microsoft Corporation:** Deploy Python Applications on Azure App Service. Available at: https://learn.microsoft.com/azure/app-service/quickstart-python

- **scikit-learn Developers:** scikit-learn: Machine Learning in Python — Official Documentation. Available at: https://scikit-learn.org/stable/documentation.html

- **scikit-learn Developers:** Model Persistence and Version Compatibility in scikit-learn. Available at: https://scikit-learn.org/stable/model_persistence.html

Microsoft Elevate

REFERENCES:

- **Google LLC:** Google Gemini API Documentation (google-genai SDK).
  Available at: https://ai.google.dev

- **Google LLC:** Gemini Models and Generate Content API Reference.
  Available at: https://ai.google.dev/docs

- **Streamlit Inc:** Streamlit Documentation — Building Data Apps in Python.
  Available at: https://docs.streamlit.io

- **Kaggle Community:** Employee Attrition Prediction using Machine Learning.
  Available at: https://www.kaggle.com

REFERENCES:

- **GitHub Link : Click Here**
- **Deployed Web URL:**https://github.com/Sruthilayaaa-K/ATTRISENSEAI---DATA-DRIVEN-INSIGHTS-FOR-EMPLOYEE-RETENTION

Microsoft Elevate

Thank You