**Software Requirements Specification (SRS)**

**Background:**

This SRS outlines the software requirements for two projects: "Calculation for Year of Graduation" and "Prediction of Placement Status." These projects involve data preprocessing, machine learning model development, and report generation to assist educational institutions in optimizing student support and placement predictions.

**Project Overview:**

**The projects aim to achieve the following objectives:**

**Calculation for Year of Graduation:**

Calculate the expected year of graduation for students based on their academic year.

Generate an Excel report ("Calculation for Year of Graduation.xlsx") containing the calculated graduation years.

**Prediction of Placement Status:**

Develop a machine learning model to predict student placement status.

Automate data preprocessing for improved data quality.

Generate an Excel report ("Prediction of Placement Status.xlsx") containing placement status predictions.

**Hardware Requirements:**

Adequate computing resources with sufficient memory and processing power for data preprocessing, model training, and report generation.

**Software Requirements:**

**Data Analysis and Preprocessing:**

Python (3.7+)

pandas

scikit-learn

openpyxl

**Machine Learning Model Development:**

scikit-learn

RandomForestClassifier for placement status prediction.

Data preprocessing using SimpleImputer and StandardScaler.

**Report Generation:**

Excel report generation using openpyxl.

**Constraints:**

Data privacy and security regulations must be followed throughout the project.

Models should be developed using scikit-learn.

The project timeline is limited to six weeks for completion.

**Assumptions:**

Adequate historical student data is available for both projects.

Required support and resources will be available throughout the project.

Model deployment will adhere to data privacy and security regulations.

Timely access to required data and computing resources is ensured.

**Risks:**

Data quality issues in historical student data may affect model performance.

Achieving the target accuracy for placement status prediction could pose a challenge.

Delays in data preprocessing or model development may impact the project timeline.

Availability of computing resources and software licenses may be limited.

**Functional Requirements:**

**Data Preprocessing:**

Load historical student data.

Handle missing data using SimpleImputer.

Normalize numerical data using StandardScaler.

Encode categorical data using LabelEncoder.

**Graduation Year Calculation:**

Calculate graduation years based on academic years.

**Placement Status Prediction:**

Develop a machine learning model (RandomForestClassifier) for placement status prediction.

Train the model using preprocessed data.

Predict placement status for test data.

**Report Generation:**

Generate Excel reports for both projects containing the required data.

**Non-functional Requirements:**

The machine learning model for placement status prediction should achieve a minimum accuracy of 80%.

The graduation year calculation should be precise.

Reports should be generated efficiently and with proper formatting.

**External Interface Requirements:**

**Data input:** Historical student data in Excel format.

**Data output:** Excel reports ("Calculation for Year of Graduation.xlsx" and "Prediction of Placement Status.xlsx").

**Technology Used:**

Python for data analysis, preprocessing, and model development.

scikit-learn for machine learning model development.

openpyxl for Excel report generation.