**Predicting SPY Historical Prices with Recurrent Neural Networks (RNN)**

**Task Description:**

**Objective:** The objective of this student task is to develop a deep learning model using Recurrent Neural Networks (RNN) to predict historical prices of SPY (SPDR S&P 500 ETF Trust). SPY is a widely traded exchange-traded fund that tracks the performance of the S&P 500 index.

**Task Overview:** In this task, you will use historical SPY price data available in a CSV file to create an RNN-based predictive model. This task provides hands-on experience with deep learning and time series forecasting using neural networks.

**Steps to Complete the Task:**

**1. Data Acquisition:**

* Download the historical SPY price data in CSV format.
* You can also use free financial data sources like Yahoo Finance or other reputable sources.

**2. Data Preprocessing:**

* Load the CSV file into a suitable Python environment (e.g., Jupyter Notebook).
* Explore the dataset to understand its structure, columns, and any missing values.
* Preprocess the data by handling missing values, if any, and converting date columns to a datetime format.
* Normalize the data to ensure it is suitable for RNN training.

**3. Data Sequencing:**

* Transform the time series data into sequences. Each sequence should contain a set of historical prices and the corresponding target price for prediction.

**4. Model Architecture:**

* Design an RNN model for time series forecasting. You need to use TensorFlow.
* Experiment with different types of RNNs, such as LSTM (Long Short-Term Memory) or GRU (Gated Recurrent Unit), and different architectural configurations.

**5. Model Training:**

* Train your RNN model using the training dataset. Adjust hyperparameters as needed to optimize model performance.
* Implement early stopping to prevent overfitting.

**6. Model Evaluation:**

* Evaluate your RNN model's performance using appropriate time series metrics, such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE).
* Visualize the model's predictions against the actual SPY prices to assess its accuracy.

**7. Documentation:**

* Document your code, including comments explaining key steps and decisions.

**8. Outcome:**

* Original or Updated csv file
* H5 and pkl file
* Training file of RNN, LSTM and GRU (optional) choose method with the best performance.
* Pipeline file (how the customer works with your H5 file sample code) make prediction for the next week and show the numbers and the results.
* Word (Docx) file that Create a brief report summarizing your approach, model performance, and any insights gained from the analysis.

**Resources:**

* Python programming environment (e.g., Jupyter Notebook)
* TensorFlow
* Python libraries for data analysis (e.g., Pandas), visualization (e.g., Matplotlib/Seaborn), and machine learning (e.g., Scikit-Learn)
* Historical SPY price data in CSV format
* Online tutorials and documentation for deep learning with RNNs and time series forecasting

**Task Duration:** typically takes 4-6 weeks to complete.