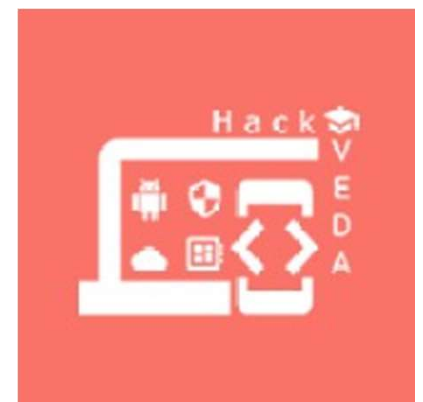




Collecting historical stock market data from Alpha Vantage

Md Firoz Alam
AWS Machine Learning Intern
17/07/24



STORY HOOKS



- ▶ Suppose Ramesh is a data analyst in Mumbai, he was working on a stock prediction project.
- ▶ He used Alpha Vantage API to collect data from the Bombay Stock Exchange (BSE) and National Stock Exchange (NSE).
- ▶ Rahul then cleaned the collected data for analysis.



Table Of Contents

- ▶ Introduction
- ▶ Different Phases
- ▶ Data Collection
- ▶ Preprocessing The Data
- ▶ Saving into CSV Format
- ▶ Demonstration
- ▶ Conclusion
- ▶ References
- ▶ Q&A Session

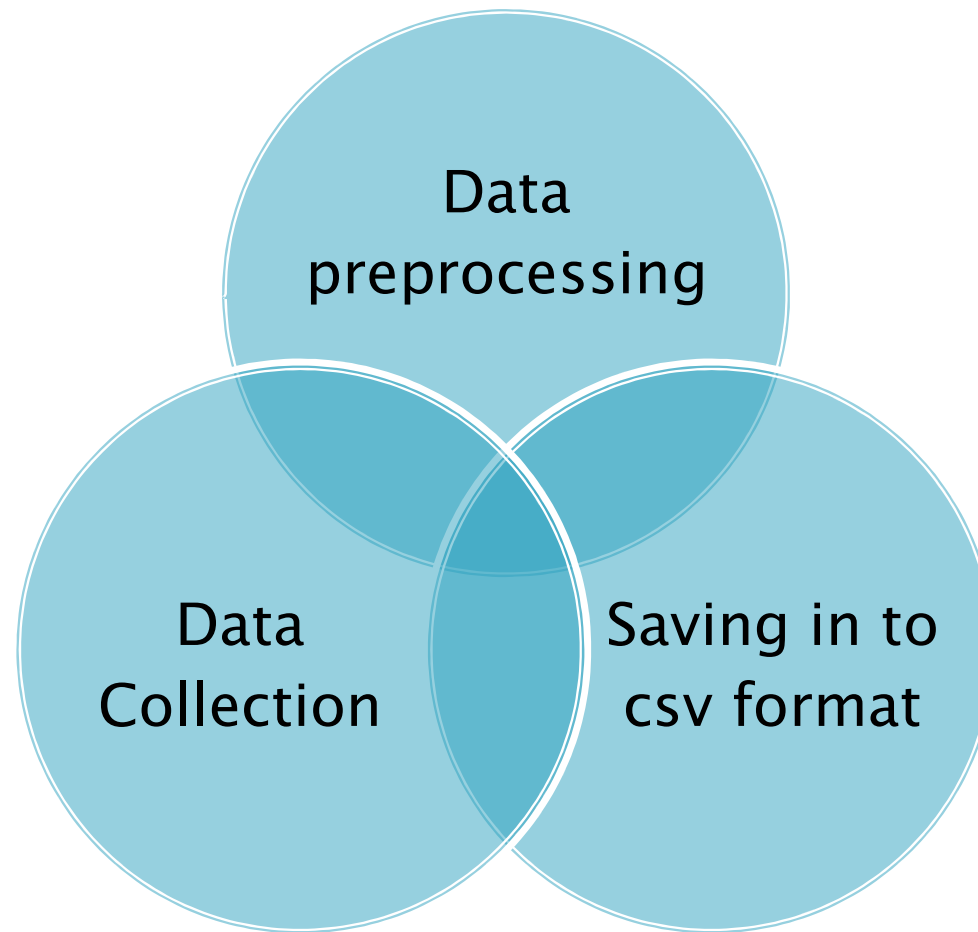


Introduction

- ▶ Alpha Vantage provides real-time and historical financial data through an API, including stocks, forex, and crypto currencies. It offers detailed market data like daily prices and volume, and is widely used for financial models, market analysis, and data integration due to its accessibility and comprehensive data.



Different Phases



Data Collection

Setting Up a Free Alpha Vantage API Key:

- ▶ To access data on stocks, forex, crypto currencies, and technical indicators, you need an API key from Alpha Vantage. Here's how to get started
- ▶ **Register for a Free API Key:** Visit Alpha Vantage Support and register for a free API key.
- ▶ **Using the API Key :**After generating the key, you can use it to make API calls and access various financial data.
- ▶ **Example Use Case:** With this API key, you can fetch daily stock market data, analyze trends, and integrate this information into your financial applications for enhanced decision-making.



Preprocessing The Data

- ▶ To ensure the integrity and reliability of your dataset, it's crucial to preprocess the data by addressing any missing or inconsistent values. Here's how to do it effectively
- ▶ **Identify Missing Values:**—First, inspect your dataset for any missing values using methods like `isnull()` or `info()` in pandas.

Handling Missing Values:

- ▶ **Replace Missing Values:** Use the `fillna()` method to replace missing values with a specified value, such as the mean, median, or a constant.
- ▶ **Remove Rows with Missing Values:** If the missing values are not suitable for replacement, use the `dropna()` method to remove rows with missing data.
- ▶ **Consistent Data Formatting:** Ensure that all data entries are consistently formatted. For example, dates should follow a uniform format, and categorical data should be standardized.



Saving into CSV Format

- ▶ After cleaning and preprocessing your data, it's important to save the resulting data frames for use in subsequent analysis or modeling tasks. This ensures that you don't have to repeat the preprocessing steps, saving time and maintaining consistency. Here's how you can do it
- ▶ **Save the Data frame:** Use the `to_csv()` method in pandas to save the data frame to a CSV file.
- ▶ **Specify the Filename:** Choose a descriptive filename that reflects the contents or the stage of preprocessing.
- ▶ **Check the Saved File:** Verify the saved file to ensure it has been stored correctly and includes all necessary data.



Demonstration

- ▶ To collect relevant historical stock market data for NSE and BSE using the Alpha Vantage API
- ▶ To preprocess and save the data in CSV format for subsequent analysis and modeling



Conclusion

- We used the Alpha Vantage API to collect historical stock data for NSE and BSE indices. The data was cleaned for reliability and predictive models were developed using decision trees, logistic regression, and random forests. The processed data was saved as CSV files, facilitating efficient workflows and forming a solid base for advanced financial analysis."
- By successfully addressing the key aspects of data acquisition, preprocessing, and model development, this project demonstrates a comprehensive approach to **leveraging financial** data for predictive analytics



References

- ▶ Alpha Vantage. (n.d.). Alpha Vantage API Documentation. Retrieved from <https://www.alphavantage.co/documentation/>
- ▶ Pandas Documentation. (n.d.). Retrieved from <https://pandas.pydata.org/docs/>
- ▶ Matplotlib Documentation. (n.d.). Retrieved from <https://matplotlib.org/stable/contents.html>
- ▶ You Tube
- ▶ Hack Veda References Video



Q&A Session





firozalam1160006@gmail.com