**INTERACTIVE DATA**

**FINAL PROJECT REORT**

**-Teja Tarapatla**

**Stock Market Prediction using CNN-based Modeling**

**Purpose**

In terms of stock market patterns, this research shows how interactive data visualizations may help financial experts make data-driven judgments. Using Convolutional Neural Networks (CNNs) to forecast stock prices and display the results on an interactive dashboard, the project seeks to provide an answer to the following query:

*How well can CNN-based models forecast trends in the stock market, and how can these forecasts be represented to help financial analysts make decisions?*

The problem is carefully framed in this project, which guides the creation of an easy-to-use, interactive tool to help analysts comprehend predictions and make wise decisions.

**Overview and Driving Question**

The main question driving for this project is:  
*With CNN-based predictions, what trends and insights can be obtained to enhance comprehension of stock market movements?*

The project enables financial analysts to connect CNN predictions to a methodically organized data visualization approach to:

1. Compare actual vs. predicted stock prices.
2. Identify trends and patterns over time.
3. Analyze prediction accuracy using interactive metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE).

**Data Sources, Quality, and Structure**

1. **Data Sources**:
   * Historical stock price datasets, including NASDAQ, NYSE, DJI, and Russell indices.
   * The daily open, high, low, close, and trading volumes are among the features.
2. **Data Quality**:
   * Using mean imputation, the data was preprocessed to deal with missing values.
   * MinMaxScaler was used to normalize features to guarantee consistency and suitability for CNN input specifications.
3. **Data Structure**:
   * A 3D structure (samples, time steps, and features) was created from the preprocessed data to meet the CNN model's specifications.
   * For visualization, the dataset contains fields for Date, Stock Symbol, Actual Price, and Predicted Price.

**Process: Collection, Wrangling, Analysis, and Interactive Development**

1. **Collection and Wrangling**:
   * To deal with missing values, the data was cleaned, and feature ranges were normalized through scaling.
   * To make analysis easier, stock indices were combined into a single structured format.
2. **Analysis and Modeling**:
   * Convolutional layers for feature extraction, pooling layers for dimensionality reduction, and dense layers for predictions were all included in the CNN architecture.
   * Using the Adam optimizer and Mean Squared Error (MSE) as the loss function, the model was trained and demonstrated consistent convergence across training epochs.

A graph of training and validation loss

Description automatically generated

1. **Interactive Visualization**:
   * Using Dash and Plotly, the dashboard was created to give users the following:
     + Dropdown menus for choosing a stock symbol.
     + Date range sliders for time filtering.
     + Interactive line charts show the difference between the actual and expected stock prices.
     + Residual plots to show prediction errors.
     + The MAE and MSE summary metrics are dynamically updated in response to user selections.

A screen shot of a graph

Description automatically generated

**Application of Course Concepts**

* **Data Analysis**: Effective preprocessing ensured data was clean and ready for deep learning and visualization.
* **Machine Learning**: By utilizing the fundamental ideas of convolution and pooling layers, a CNN model was customized for time-series forecasting.
* **Visualization Design**: Clarity, user interaction, and relevance to financial analysts decision-making processes were given top priority in the dashboard design.

**Audience-Centric Design and Development**

The project's main target audience was financial analysts. To prioritize their needs:

1. **Intuitive Interaction**: Stock trend and prediction accuracy can be easily explored with dropdown menus, sliders, and dynamic plots.
2. **Relevant Metrics**: For reliability assessment, key performance metrics such as MAE and MSE are shown prominently.
3. **Actionable Insights**: Analysts can spot inconsistencies and patterns with the help of visually represented forecasts and patterns.

**Reflections: Limitations and Future Directions**

1. **Limitations**:
   * Without considering outside variables like news sentiment or macroeconomic indicators, the model only uses historical price and volume data.
   * Because it is limited to historical datasets, the project's real-time predictive capabilities are limited.
2. **Future Directions**:
   * **Integration of External Factors**: To make predictions more accurate, include macroeconomic indicators or news sentiment.
   * **Real-Time Prediction**: Utilizing live data APIs, modifying the dashboard to make predictions in real time.
   * **Advanced Visualizations**: Add correlation matrices or heatmaps to the dashboard to conduct more in-depth multi-index analyses.

**Final Reflections**

This research successfully created an interactive application that efficiently combines sophisticated machine learning with user-friendly data visualization, allowing financial analysts to extract insights from stock projections. CNN-based modeling in conjunction with interactive dashboards accomplishes the objectives of the course and establishes a basis for future growth.