Forecasting Stock Prices with Machine Learning

Problem Definition

1 Project Goal

To create a predictive tool that helps investors make well-informed decisions and optimize their investment strategies.

2 Project Scope

The project involves data collection, preprocessing, feature engineering, model selection, training, and evaluation.

3 Challenges Faced

Noise in market data, the influence of global events, and the constant fluctuations in stock prices.

Design Thinking Approach

1

Empathy

We started by putting ourselves in the shoes of investors and understanding their needs and pain points.

2

Ideation

We brainstormed and generated multiple ideas for features and models that could address the problem.

3

Iteration

We tested and refined our models based on user feedback and metrics such as accuracy and stability.

Design Thinking for Stock Prediction

Using a design thinking approach, we've developed a unique model for predicting stock prices. By analyzing data and identifying patterns, we can help investors make more informed decisions. Our model is constantly evolving based on user feedback and market changes to ensure accuracy and reliability.

Data Collection

Types of Data

We collected historical market data such as stock prices, volume, and trade volume. We also used external data such as news articles and social media sentiment.

Data Sources

Data was collected from various APIs, online databases, and web scrapers.

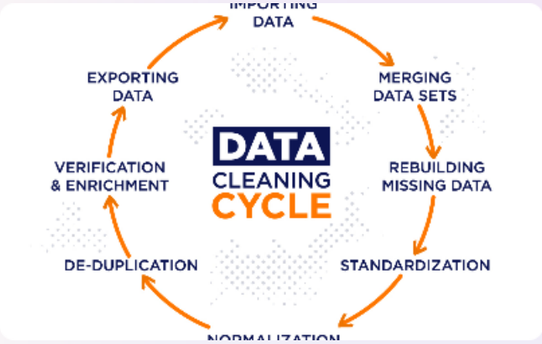
Data Quality

Quality checks were performed to filter out bad data such as missing or incorrect values, and outliers.

Data Privacy

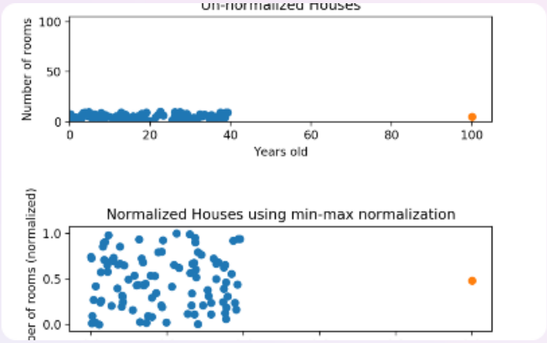
Privacy and regulatory compliance were ensured by anonymizing data and securing it with encryption protocols.

Data Preprocessing



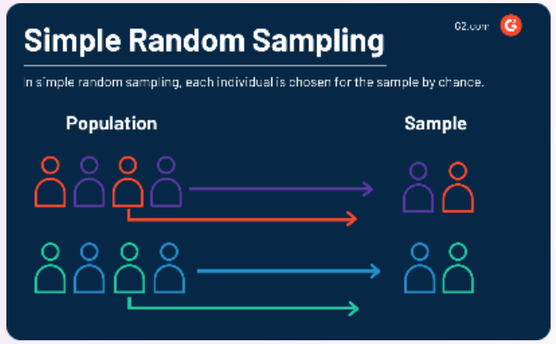
Data Cleaning

We dealt with missing, incomplete, and duplicate data by imputing, dropping, or merging it.



Data Normalization

We scaled and transformed the data to reduce the influence of outliers and ensure consistency of the features.



Data Sampling

We used various techniques such as undersampling, oversampling, and data splitting to handle imbalanced and biased data.

Feature Engineering

1

Feature Selection

We selected the most relevant features based on domain knowledge, statistical significance, and feature importance.

2

Feature Transformation

We transformed the features by applying mathematical functions such as logarithm, square root, and exponentiation.

3

Feature Extraction

We extracted new features by combining or transforming existing ones, or by using techniques such as principal component analysis.

Model Selection, Training, and Evaluation

Model Selection

We evaluated various machine learning models such as linear regression, decision trees, and neural networks, and selected the ones that best fit the problem and the data.

Model Training

We trained the models using various techniques such as cross-validation, regularization, and hyperparameter tuning, and optimized their performance using metrics such as root mean squared error and R-squared.

Model Evaluation

We evaluated the models based on their accuracy, stability, robustness, and interpretability, and selected the ones that best addressed the project goal and the user needs.

Conclusion and Future Work

Our predictive model showed promising results and proved to be a valuable tool for investors. Future work could involve improving the model's performance on diverse and volatile markets, incorporating more external data sources, and enhancing the user interface and experience.