Machine Learning for Stock Direction Prediction

A Project Implementing an ML Model with a Real-World Dataset

Prepared by: Abeer AlShehri, Abeer Alqarni

Dr. Abrar Wafa

Project Overview & Objectives

📈 Project Overview

Developed a machine learning system to forecast daily stock price direction by integrating financial news sentiment and technical market indicators.

📋 Project Scope

✅ Developed as a prototype for predictive modeling

✅ Not intended to function as an operational trading system

⚠️ Key Constraints

❌ **Data Limitation:** Absence of timestamps in news data prevented temporal correlation with stock prices

❌ **Dataset Size:** Final dataset of 1255 samples was relatively small

◎ Primary Goal

📈

Construct and evaluate a predictive model capable of forecasting whether a stock's closing price would rise or fall on the subsequent trading day.

⚙️ Key Approaches

- Technical indicators from historical price and volume data
- Sentiment indices based on financial news

System Diagram and Models Used

⚙️ Algorithm Selection

Logistic Regression
✓ Baseline model

Support Vector Machine
✓ Non-linear patterns

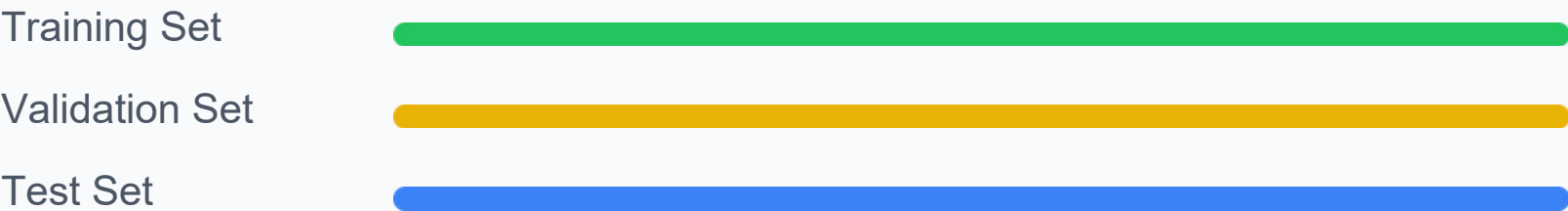### XGBoost
✓ Ensemble learning

LSTM
✓ Sequential data

📈 Robust Evaluation Approach

Chronological Data Split
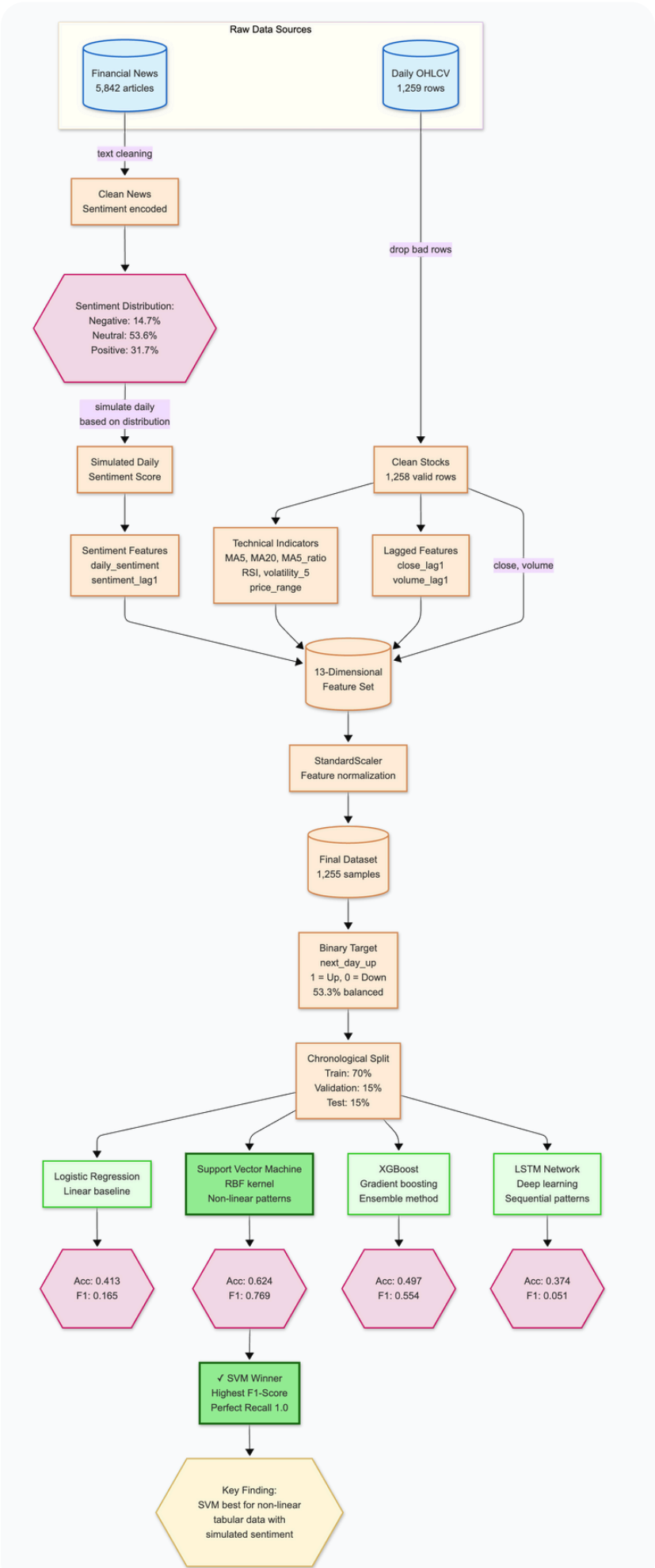
Training Set
Validation Set
Test Set

70% Training   **15%** Validation   **15%** Test

Features

Trend Indicators
- ma5
- ma20
- ma5_ratio

📈 Price & Volume Basics
- close
- volume
- price_range

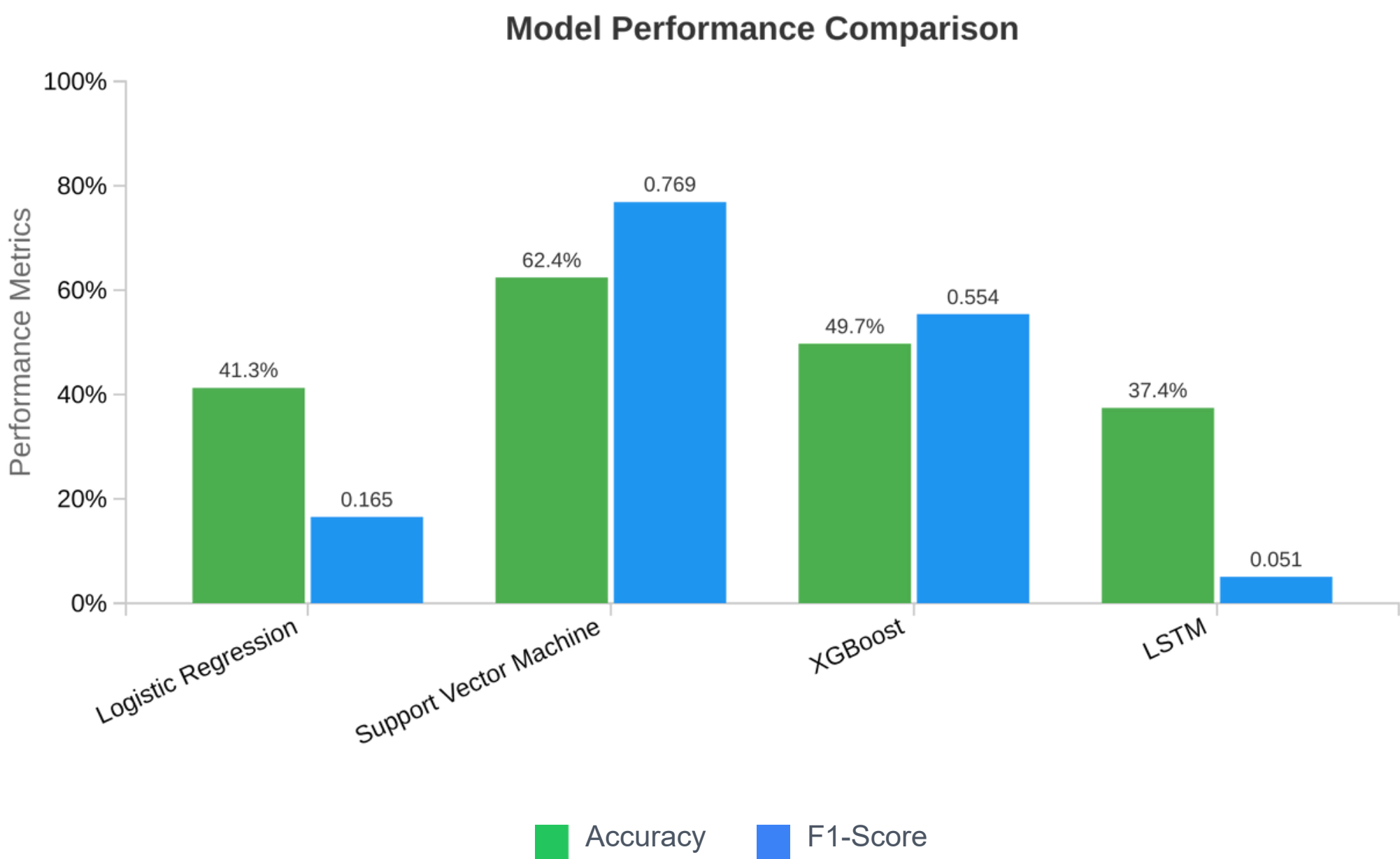### 🚀 Momentum & Volatility
- rsi
- volatility_5

🕘 Temporal Memory
- close_lag1
- volume_lag1

💬 Sentiment Features
- daily_sentiment
- sentiment_lag1

Findings & Results


Model Performance Comparison

🏆 **SVM Emerges as Top Performer**

Support Vector Machine achieved **62.4%** accuracy and **0.769** F1-Score

💡 **Key Insights**

✅ **SVM's Strength:** High recall value of 1.0 correctly identified all "up" days

⚠️ **LSTM's Weakness:** Poor performance with F1-Score of 0.0508

⚖️ **XGBoost's Balance:** F1-Score of 0.5540 showing balanced precision and recall

❌ **Logistic Regression:** F1-Score of 0.1654 confirming non-linear relationships

Critical Project Limitations

Despite successful model implementation, several critical limitations affected the project's outcomes:
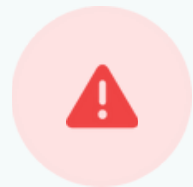
Simulated Sentiment Data

Reliance on simulated news sentiment rather than genuine, time-stamped data. This limits real-world predictive power.
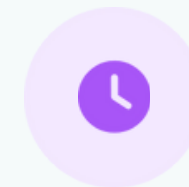
Single Stock Focus

Model developed exclusively on data from one stock, limiting generalizability across broader markets.

Basic Feature Set

Relatively simple feature engineering. More sophisticated technical indicators could enhance performance.

Static Sentiment Simulation

Simulation didn't account for dynamic relationships between sentiment spikes and market volatility.

ℹ️ Impact Assessment

These limitations collectively restrict the model's real-world applicability. Addressing them would require acquiring better data sources and expanding the feature set.

Future Work Recommendations

Key improvements to enhance the stock direction prediction system:

Timestamped News Data

Acquire news articles with exact publication dates for accurate sentiment correlation with stock movements

Expanded Dataset

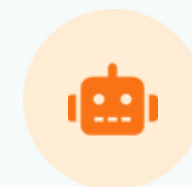Utilize 5-10 years of data on multiple stocks for a larger, more diverse dataset

Advanced Features

Explore sophisticated technical indicators and macroeconomic data

Ensemble Modeling

Develop an ensemble of best-performing models to combine their individual strengths

Backtesting Framework

Develop a comprehensive financial backtesting framework that models actual trades, including realistic transaction costs