Overview

- Introduction
- Motivation
- Literary Review
- Problem Statement
- Objectives
- Methodology
- Tools And Technology Used
- Algorithm Used
- Future Work
- Results
- Outcome and Impact

Introduction

FinSentix is an innovative platform in finance that uses machine learning and NLP to automatically extract vital insights from complex financial scripts, simplifying data access.

Motivation

Data Deluge

The financial industry is experiencing an unprecedented influx of data. Each day, terabytes of financial information are generated across markets, corporations, and economic indicators.

Demands for Innovation

In a rapidly evolving financial landscape, innovation is not just a choice; it's a necessity. The industry is ripe for transformative solutions that streamline data processing.

Time Sensitivity

In financial decision-making, every second counts. The efficiency of data processing directly impacts the speed at which informed decisions can be made.

Efficiency as a Priority

Efficiency is the lifeblood of the financial industry. Every resource saved in data processing can be redirected toward higher-value financial analysis and strategic decision-making.

Literature review

	Financial Named Entity Recognition (FNER)	NLP Models for Financial Text:	Sentiment Analysis in Finance	Key Entity Detection:	Weak Supervision in NER
AckNER	The paper introduces AckNER, a tool designed for extracting financial information from research articles and dissertations.	AckNER employs advanced algorithms like Bidirectional LSTM-CRF for sequence labeling in financial texts.	Not explicitly discussed.	Not explicitly discussed.	The paper primarily focuses on the development of AckNER for financial NER.
Sentiment analysis and key	Addresses FNER indirectly through key entity detection in online financial texts.	Applies BERT and RoBERTa models for sentiment analysis and key entity detection in financial texts.	Focuses on negative sentiment information extraction in online financial texts using RoBERTa and ensemble learning.	Addresses key entity detection in online financial texts using RoBERTa and ensemble learning.	Weak supervision is not a primary focus of this paper.
FinBERT	Primarily focuses on developing FinBERT, a pre-trained model for financial text mining.	Introduces FinBERT as a domain-specific pre-trained language model for financial text mining.	Demonstrates FinBERT's effectiveness in sentiment analysis tasks on financial data.	While not the primary focus, FinBERT can be used for key entity detection.	Weak supervision is not a primary focus of this paper.
FiNER: Financial Named Entity Recognition Dataset and Weak-Supervision Model	Focuses on improving NER accuracy in financial news articles using FiNER-ORD dataset and the FiNER weak-supervision model.	Introduces custom labeling functions (FiNER-LFs) and Snorkel for weak supervision in NER tasks within financial text.	Sentiment analysis is not the primary focus of this paper.	While not the primary focus, FiNER can be used for key entity detection.	The paper introduces FiNER as a weak-supervision framework utilizing custom labeling functions and Snorkel for label aggregation.

Literature review

	Custom Labeling Functions:	Label Aggregation Techniques	Financial Corpora and Datasets:	Performance Metrics:	Challenges in Financial NER:
AckNER	Custom labeling functions are used to enhance NER accuracy.	Label aggregation techniques are not a central part of this paper.	The paper utilizes research articles and dissertations for financial information extraction.	Performance metrics are not explicitly discussed but can be inferred based on the focus on accuracy.	The paper addresses the challenges of financial NER, including the need for specialized tools.
Sentiment analysis and key	Custom labeling functions are not explicitly discussed.	Label aggregation techniques are not a central part of this paper.	Utilizes online financial texts for sentiment and key entity detection.	Evaluates performance using accuracy and F1 score.	Addresses challenges in financial NER, including variations in naming conventions and unconventional expression of financial entities.
FinBERT	Custom labeling functions are not explicitly discussed.	Label aggregation techniques are not a central part of this paper.	Utilizes financial text data for pre-training FinBERT.	Evaluates FinBERT's performance in tasks like sentiment analysis and question answering.	Addresses the challenge of limited labeled data in the financial domain.
FiNER: Financial Named Entity Recognition Dataset and Weak-Supervision Model	Introduces FiNER-LFs, custom labeling functions designed for financial NER.	Utilizes Snorkel's weighted majority vote aggregation for label aggregation.	Creates the FiNER-ORD dataset, specialized for financial NER.	Evaluates NER model performance using F1-score and weighted average F1-score.	Addresses challenges in financial NER, including difficulty in handling overlapping entities.

Problem Statement

- The financial sector is grappling with the increasing volume and complexity of textual financial data, requiring manual categorization methods that face challenges like time-consuming processes and poor information retrieval.
- Machine learning and NLP technologies automate textual financial data processing, reducing errors and time, enabling more effective analysis and informed decision-making in the changing financial landscape.

Objectives

Objective 01

To Automate Numerical Entity Recognition and Improve Labelling Accuracy

Objective 02

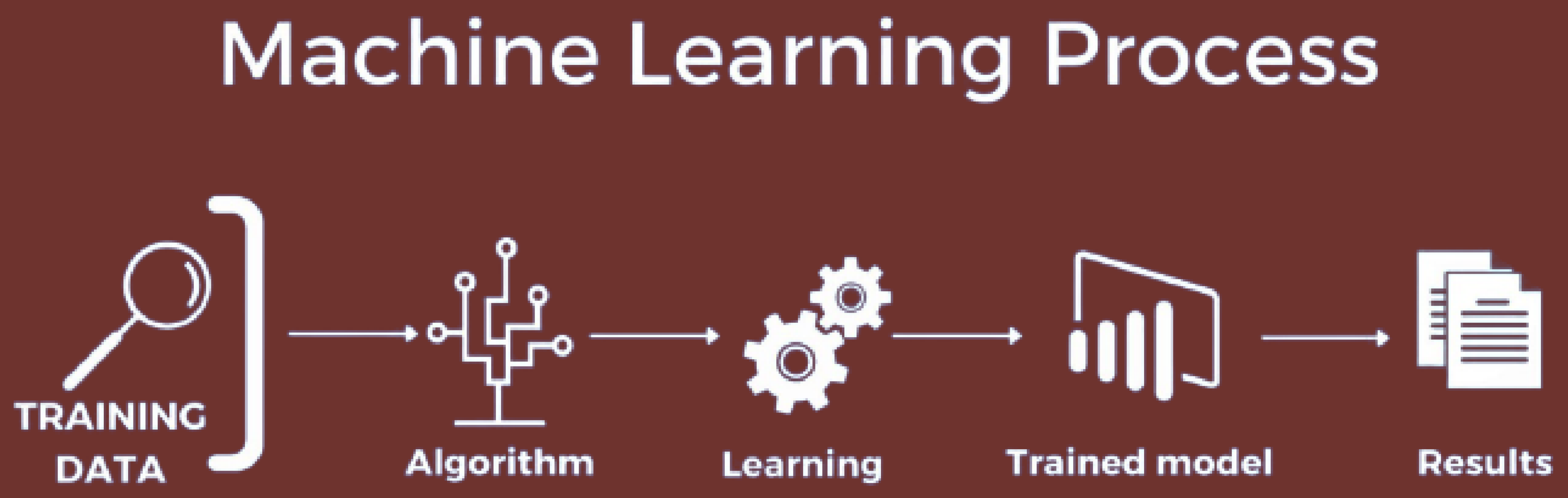
To Enrich Contextual Understanding with Optimise Efficiency

Objective 03

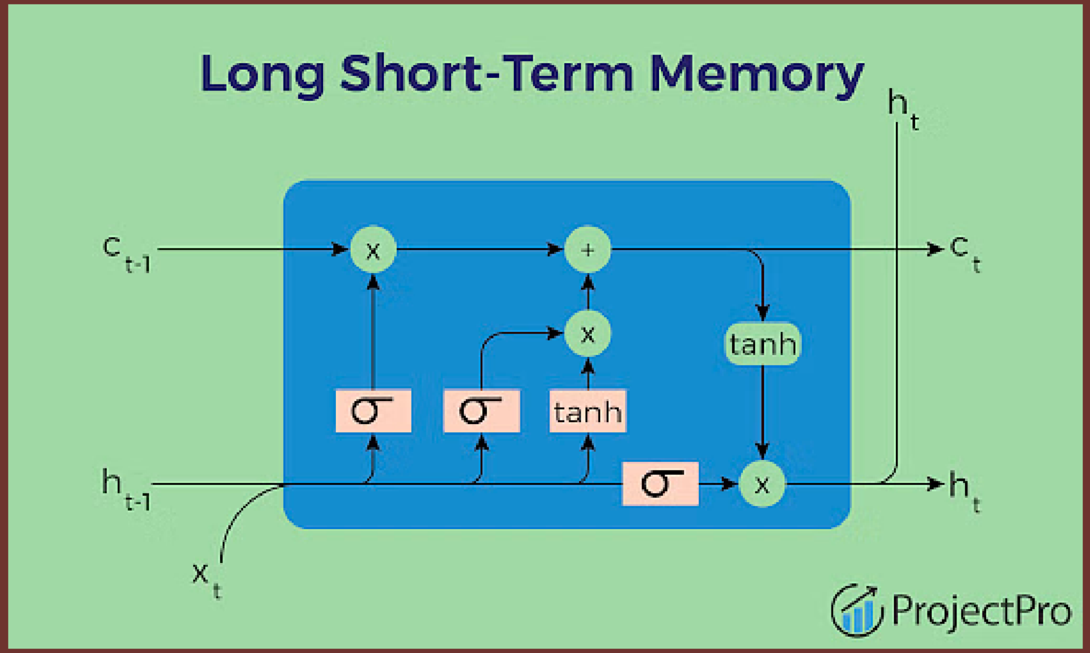
To Provide Real-time Insights

Methodology

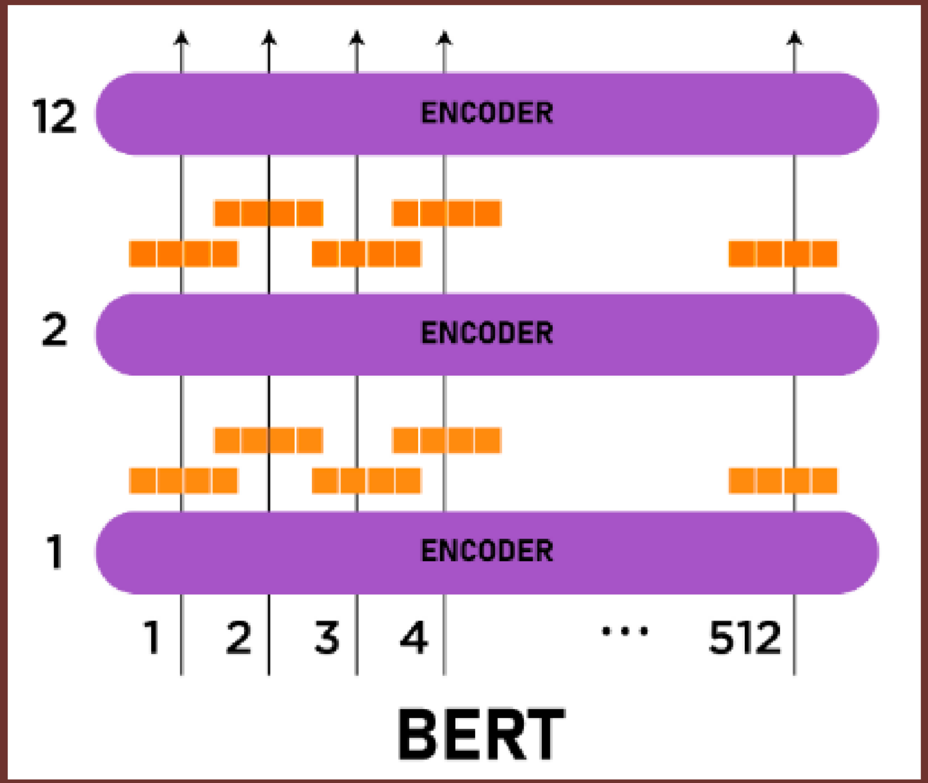
1. Data Collection and Preprocessing
2. NLP Model Selection and Customization
3. Labeling and Annotation
4. Training and Validation
5. ML Model Training
6. Model Evaluation and Real-time Prediction



Technology and Tools Used



LSTM (Long Short-Term Memory) is a type of recurrent neural network designed for handling long-term dependencies in data sequences.
LSTMs excel at preserving and utilizing context information over extended sequences in tasks like speech recognition and language modeling.

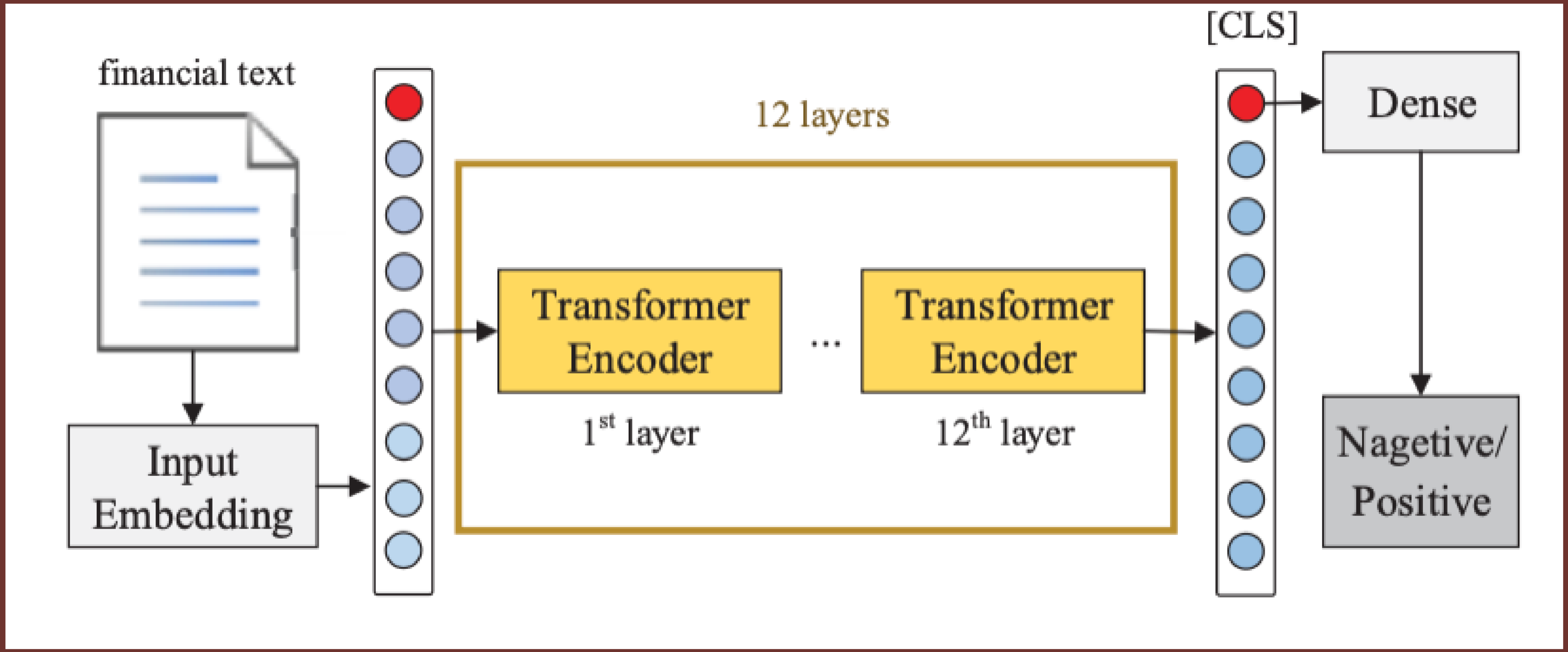


BERT (Bidirectional Encoder Representations from Transformers) is a transformer-based model for natural language processing, understanding context in both directions for various language tasks.

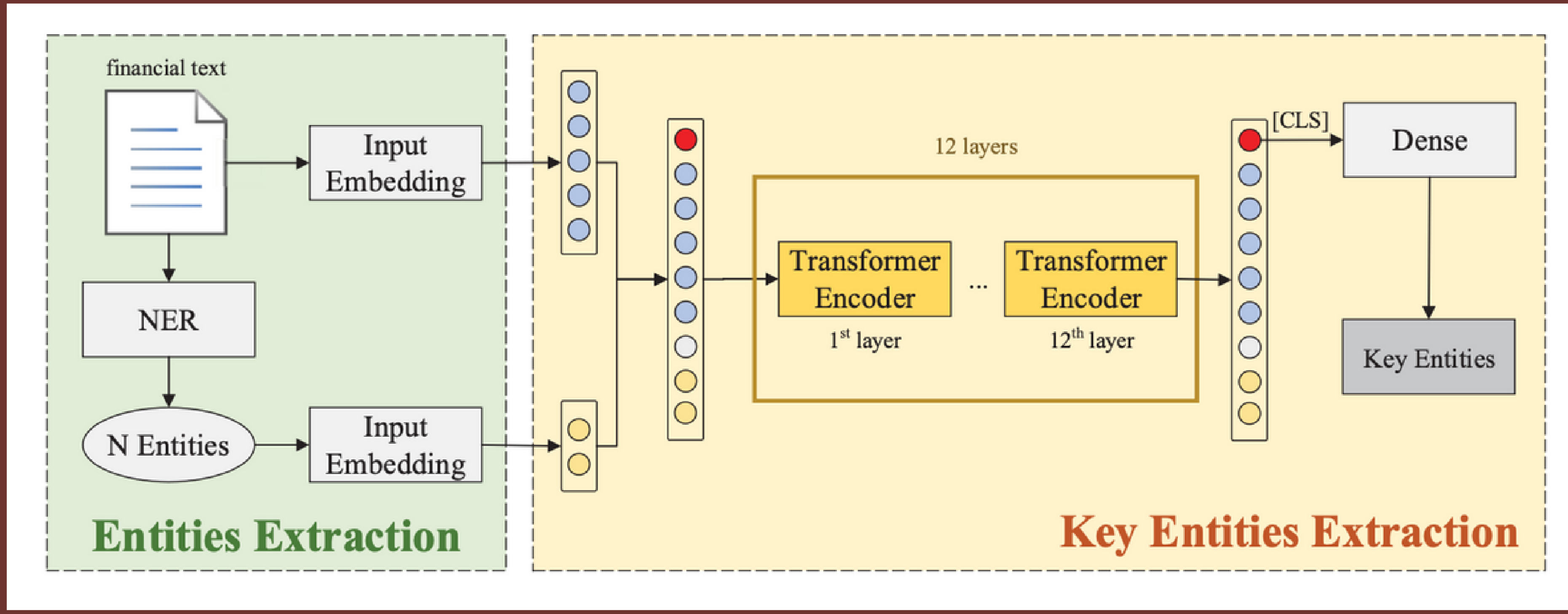
Python is a versatile, high-level programming language known for its simplicity, readability, and wide range of applications in software development and data analysis.

Algorithm Used

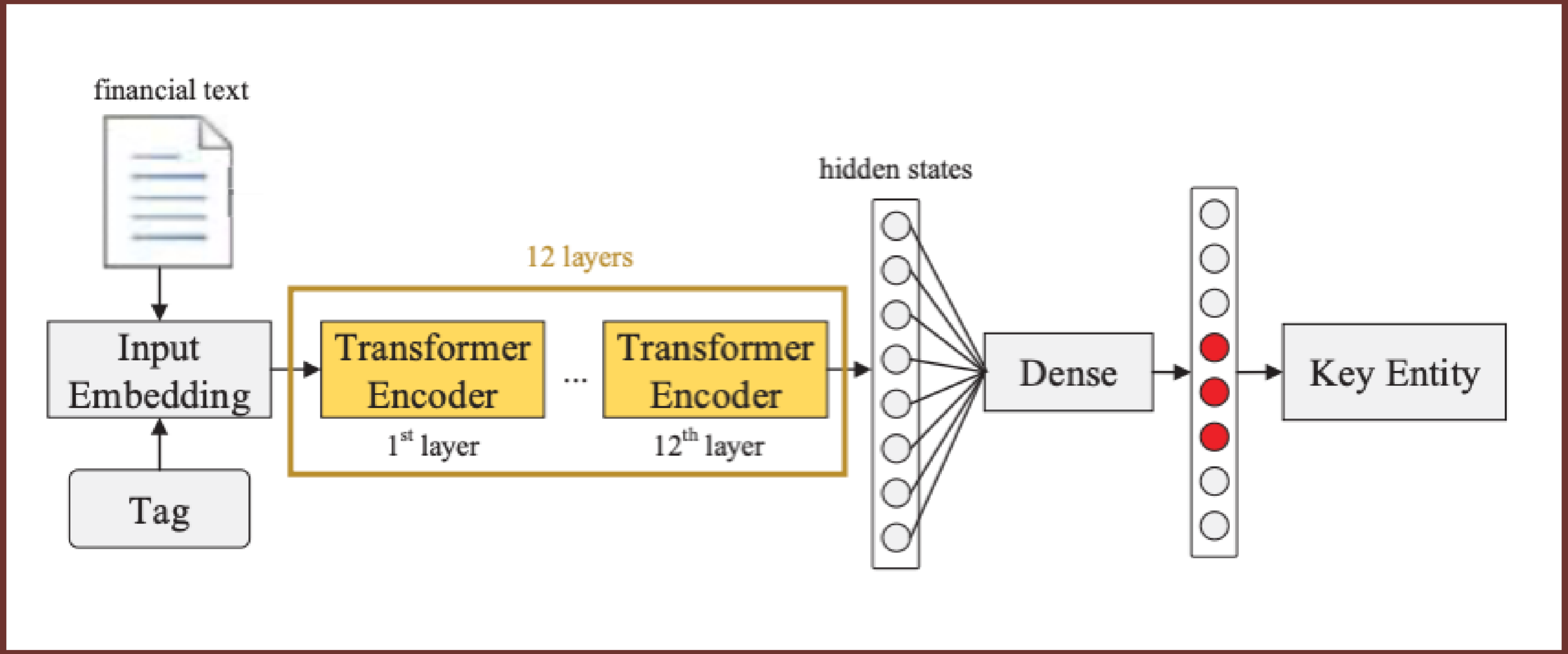
1) Transform financial text via a transformer-based model to predict sentiment—whether positive, negative, or neutral—regarding financial news.



2) BERT tags and labels financial entities, values, and keywords, enabling extraction of structured data for easier analysis and decision-making in financial content.



3) Extract key financial entities associated with identified tags to enhance understanding and analysis of relevant sentiments and traits in financial language.

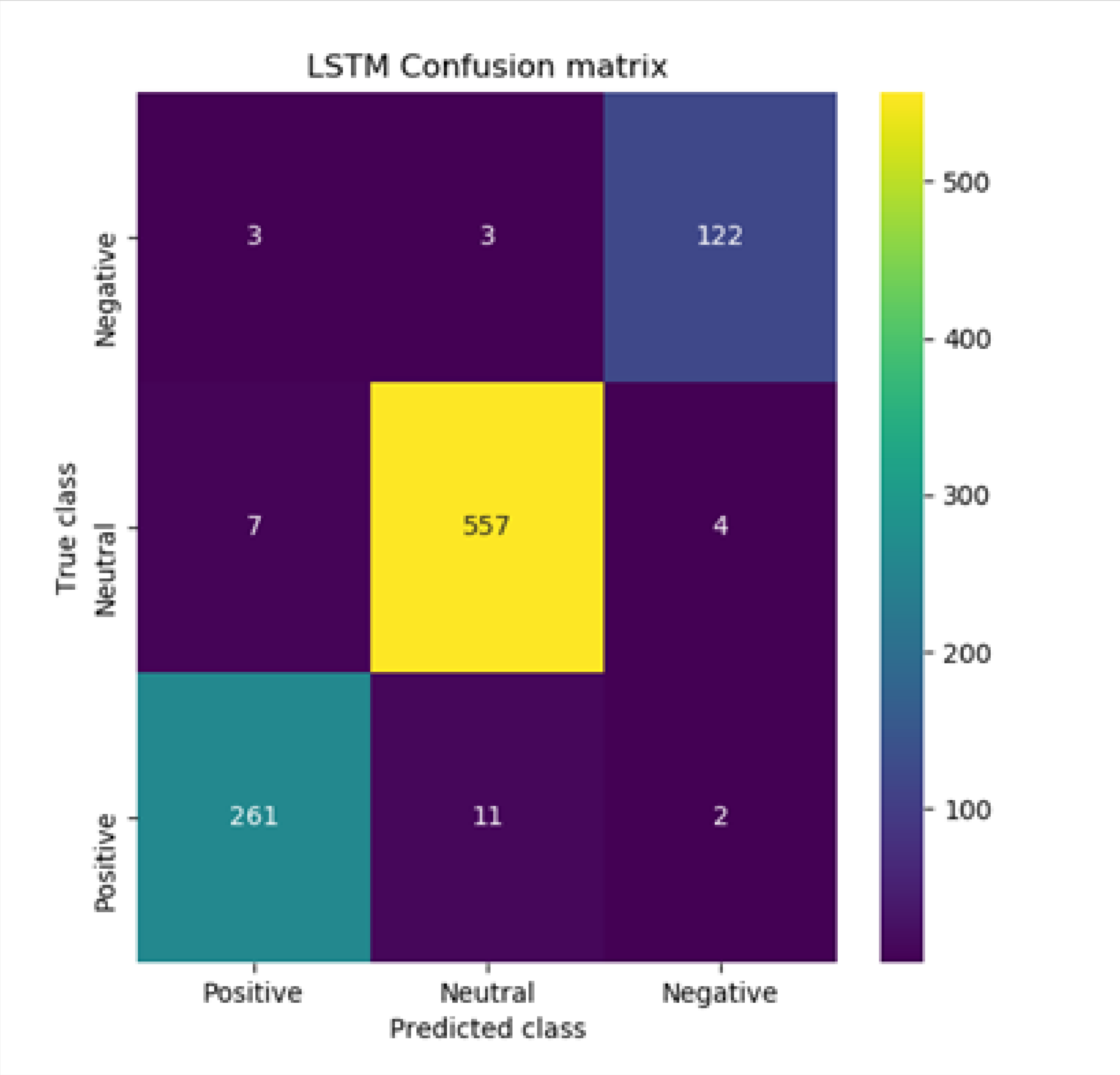


Results

Confusion Matrix

LSTM got 96% accuracy in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels

LSTM Accuracy : 96.907216494S4536
LSTM Precision : 96.39020807349054
LSTM Recall : 96.2104SIS7SOS163
LSTM FScore : 96.29913623329655



Validation Dataset

Displaying NEWS text from TEST data and then displaying extracted numeric %, prices, sentiments and sentiments predicted scores and this prediction you can see for each NEWS text line

```
Financial News = ['Technopolis plans to develop in stages an area of no less than 100,000 square meters in order to host companies working in computer technologies an
d telecommunications , the statement said .']
Extracted % = []
Extracted Prices = ['100', '000']
Predicted Sentiments = Neutral
Sentiments Score = 1.0

Financial News = ['Net sales increased to EUR193.3 m from EUR179.9 m and pretax profit rose by 34.2% to EUR43.1 m. ( EUR1 = USD 1.4 )']
Extracted % = ['193.3', '179.9', '34.2', '43.1', '1.4']
Extracted Prices = ['1']
Predicted Sentiments = Positive
Sentiments Score = 0.9999995

Financial News = ['Net sales surged by 18.5% to EUR167 .8 m. Teleste said that EUR20 .4 m , or 12.2% , of the sales came from the acquisitions made in 2009 .']
Extracted % = ['18.5', '12.2']
Extracted Prices = ['167', '.8', '20', '.4', '2009']
Predicted Sentiments = Positive
Sentiments Score = 0.9997476

Financial News = ["The fuel purchase contracts have been signed with three months ' delivery from this September to November ."]
Extracted % = []
Extracted Prices = []
```

Future Work

1. Enhanced Data Enrichment
2. Real-time Integration
3. Machine Learning Refinement
4. Integration with Financial Tools
5. Scalability
6. Documentation and Support
7. Feedback Mechanism
8. Ethical Considerations

Research Outcome and Impact

- Advanced Credit Risk Assessment
- Market Trend Prediction
- Fraud Detection and Prevention
- Algorithmic Trading Strategies
- Customer Behavior Analysis
- Regulatory Compliance Solutions
- Risk Management Frameworks
- Economic Forecasting