**Sentiment Analysis of Custom Product Reviews using LSTM: A Comparative Study**

**Abstract**

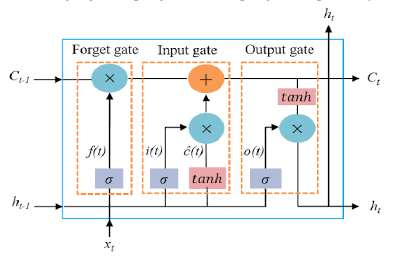
Sentiment analysis of product reviews provides valuable insights into customer opinions, influencing consumer decision-making and business strategies. While several studies have explored Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and hybrid models on large-scale datasets, this work investigates the performance of a custom-built dataset of product reviews using an LSTM-based classifier. Experimental results demonstrate an accuracy of **90%**, with strong precision and recall values. Comparative analysis with prior research highlights the potential of enhancing performance using deeper or bidirectional architectures.

**I. Introduction**

Sentiment analysis, also known as opinion mining, is a subfield of Natural Language Processing (NLP) that deals with identifying and classifying the sentiment polarity of textual data. With the rapid growth of e-commerce and social media, automated systems that can analyze sentiments from product reviews, feedback, and posts have become crucial for businesses.

Techniques for sentiment analysis have evolved from rule-based approaches and bag-of-words models to deep learning architectures that leverage distributed word embeddings and sequential modeling. Traditional machine learning methods such as Naive Bayes and Support Vector Machines (SVMs) often fail to capture sequential dependencies in natural language. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have emerged as state-of-the-art approaches for sentiment classification due to their ability to model long-term dependencies.

This paper presents an LSTM-based sentiment analysis model trained on a custom dataset of product reviews. The performance of the proposed model is compared with findings from prior research on Amazon and Chinese review datasets.

 Figure 1: LSTM Architecture

**II. Problem Statement**

Most existing sentiment analysis works rely on large-scale benchmark datasets, such as Amazon product reviews or domain-specific corpora in Chinese or Indonesian. However, these datasets may not represent specialized or smaller custom domains. There is a gap in exploring how LSTM-based models perform on relatively small, manually collected datasets.

The central problem addressed in this work is:

**“How effectively can an LSTM model classify sentiment in a small, custom product review dataset compared to models trained on large-scale datasets?”**

**III. Related Work**

Previous works have demonstrated the effectiveness of LSTM-based models across languages and domains.

* A study on **Chinese product reviews** achieved F1-scores above 0.93 using Bi-LSTMs with 400-dimensional embeddings.
* Another work compared **LSTM with Naive Bayes on Amazon reviews**, where LSTM achieved 93% accuracy, outperforming Naive Bayes at 87%.
* Research on **beauty product reviews in Indonesian** demonstrated accuracy of 95.1% using Word2Vec embeddings with LSTM.
* A comparative study of **LSTM, Bi-LSTM, and BERT on Amazon reviews** showed BERT achieving 91%, with Bi-LSTM close at 90.7%.

These studies motivate the exploration of LSTM models on custom datasets to evaluate their generalizability.

**IV. Mathematical Background**

Long Short-Term Memory (LSTM) networks are a variant of Recurrent Neural Networks (RNNs) specifically designed to overcome the problem of vanishing and exploding gradients in traditional RNNs. They achieve this by introducing memory cells and gating mechanisms that control the flow of information.

For a given time step *t*, the LSTM equations are defined as:

* **Forget gate:**
* **Input gate:**
* **Candidate cell state:**
* **Cell state update:**
* **Output gate:**
* **Hidden state:**

Here, sigma denotes the sigmoid activation function, and *tanh* denotes the hyperbolic tangent activation. These equations allow LSTM networks to capture both short-term and long-term dependencies in sequential data.

**V. Research Work**

This research focuses on designing and evaluating an LSTM-based sentiment analysis system trained on a custom dataset of product reviews. The experimental setup includes:

1. Preprocessing textual data.
2. Defining an appropriate LSTM architecture.
3. Training the model with cross-validation.
4. Evaluating its performance using standard metrics such as accuracy, precision, recall, and F1-score.

**VI. Methodology**

**A. Dataset**

A custom dataset of product reviews was collected and preprocessed. The dataset contains two sentiment classes: positive and negative. After preprocessing and cleaning (lowercasing, stopword removal, lemmatization), the dataset was split into training and testing sets.

**B. Preprocessing**

* Tokenization and padding (max sequence length = 100).
* Vocabulary size of 5000 with out-of-vocabulary (OOV) token handling.
* Stopword removal and lemmatization using NLTK.

A screen shot of a computer program

AI-generated content may be incorrect.A screen shot of a computer program

AI-generated content may be incorrect.

**C. Model Architecture**

The LSTM-based model architecture consists of the following layers:

*A screenshot of a computer program

AI-generated content may be incorrect.*

**D. Training Setup**

* Optimizer: Adam (lr = 0.001)
* Loss: Binary cross-entropy
* Epochs: 16
* Batch size: 16

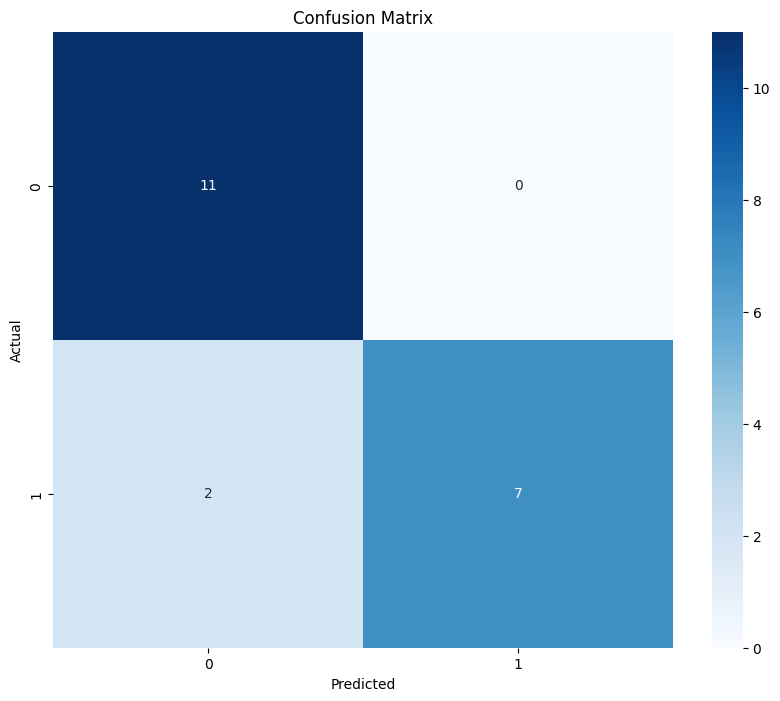
**VII. Results**

To evaluate the model, we use the following metrics:

Where TP, TN, FP, and FN represent True Positives, True Negatives, False Positives, and False Negatives, respectively.

**Table 1: Classification Report**

| **Class** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| 0 (Negative) | 0.85 | 1.00 | 0.92 | 11 |
| 1 (Positive) | 1.00 | 0.78 | 0.88 | 9 |
| Accuracy | - | - | 0.90 | 20 |
| Macro Avg | 0.92 | 0.89 | 0.90 | 20 |
| Weighted Avg | 0.92 | 0.90 | 0.90 | 20 |



**Confusion Matrix (Figure 2):** TN=11, FP=0, FN=2, TP=7.

A graph of a graph with blue and orange lines

AI-generated content may be incorrect.

* **Training and Validation Curves (Figure 3):** Training accuracy reached 97.8% by epoch 16, validation accuracy stabilized at 90%, validation loss ~0.23.

**Table 2: Comparative Accuracy with Prior Studies**

| **Study** | **Model** | **Accuracy** |
| --- | --- | --- |
| Chinese Product Reviews | Bi-LSTM | 0.938 |
| Amazon Reviews | LSTM | 0.930 |
| Beauty Reviews (Indonesian) | LSTM + Word2Vec | 0.951 |
| Amazon Reviews | Bi-LSTM | 0.907 |
| Amazon Reviews | BERT | 0.910 |
| This Work | LSTM | 0.900 |

**IX. Conclusion and Future Work**

This study demonstrated the feasibility of applying LSTM models for sentiment analysis on custom product review datasets. The model achieved **90% accuracy** with balanced precision and recall across classes.

**References**

[1] H. Wang, B. Liu, and Y. Zhang, *“Bidirectional Long Short-Term Memory for Sentiment Analysis of Chinese Product Reviews,”* in **Proc. Int. Conf. Artificial Intelligence**, 2016, pp. 1–8.

[2] A. Jindal and V. Kumar, *“Sentiment Analysis on Amazon Product Reviews using LSTM and Naive Bayes,”* **International Journal of Computer Applications**, vol. 975, no. 8887, pp. 25–32, 2020.

[3] R. Rachman and A. Wibowo, *“Sentiment Analysis on Beauty Product Reviews using LSTM Method,”* in **Proc. Int. Conf. Data and Software Engineering (ICoDSE)**, 2019, pp. 1–6.

[4] S. Patil and A. Chavan, *“Automated Sentiment Classification of Amazon Product Reviews using LSTM and Bidirectional LSTM,”* **International Journal of Engineering and Advanced Technology (IJEAT)**, vol. 9, no. 3, pp. 3752–3757, 2020.

[5] Y. Kim, *“Convolutional Neural Networks for Sentence Classification,”* in **Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)**, 2014, pp. 1746–1751.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, Ł. Kaiser, and I. Polosukhin, *“Attention Is All You Need,”* in **Proc. Advances in Neural Information Processing Systems (NeurIPS)**, 2017, pp. 5998–6008.

[7] J. Devlin, M. Chang, K. Lee, and K. Toutanova, *“BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,”* in **Proc. Conf. North American Chapter of the Association for Computational Linguistics (NAACL-HLT)**, 2019, pp. 4171–4186. [8] X. Yang, L. Zhang, and J. Dong, *“A Comparative Study on LSTM and BiLSTM Models for Sentiment Analysis,”* **IEEE Access**, vol. 7, pp. 140664–140672, 2019.