Exposé - Masterarbeit Alperen Dagli

Alperen Dagli

September 2024

1 Introduction

In recent years, blockchain technology has gained traction in various industries, offering decentralized, transparent, and secure platforms for executing transactions and processes. One of the potential applications of blockchain is in the insurance industry, where predictive models are used to assess risk, calculate premiums, and process claims. However, traditional insurance companies rely heavily on centralized infrastructures, which can be costly and time-consuming.

This thesis aims to explore the feasibility and cost-efficiency of deploying pretrained neural network models for insurance predictions on blockchain platforms. Specifically, we will compare two blockchain solutions—BNB Chain and Base Blockchain—with the conventional methods used by insurance companies for processing insurance claims and predictions.

The central hypothesis is that blockchain-based insurance predictions can offer more cost-efficient, scalable, and secure alternatives to traditional approaches, without compromising prediction accuracy.

2 Problem Statement

Traditional insurance companies spend significant resources on IT infrastructure, personnel, and operational costs to manage predictive models and process claims. These centralized processes may introduce inefficiencies, higher costs, and slower processing times. Blockchain technology, on the other hand, promises decentralized computation, reduced transaction costs, and automated smart contracts, which may lead to more efficient claim predictions.

The key question addressed in this thesis is: Can blockchain platforms such as BNB Chain and Base Blockchain offer more cost-efficient methods for making insurance predictions compared to traditional insurance companies?

3 Objectives

- **Primary Objective:** To evaluate the cost-efficiency of deploying pretrained neural network models for insurance claim predictions on blockchain platforms (BNB Chain and Base Blockchain).

- **Secondary Objective:** To compare the gas fees and transaction costs on these blockchain platforms with the operational costs incurred by traditional insurance companies for similar predictive tasks.

4 Research Questions

1. How much gas is consumed for processing insurance predictions on BNB Chain and Base Blockchain?
2. What is the cost of blockchain transactions compared to the operational costs (IT infrastructure, personnel, etc.) of traditional insurance companies?
3. Can blockchain platforms provide the same or better performance (accuracy, speed, and scalability) in making insurance predictions?
4. What are the trade-offs between blockchain-based and traditional insurance prediction methods in terms of security, transparency, and decentralization?

5 Methodology

The methodology for this thesis involves the following steps:

5.1 Blockchain Selection

The blockchain platforms chosen for this study are BNB Chain and Base Blockchain. These platforms are selected for their cost efficiency, scalability, and established infrastructure for decentralized applications.

5.2 Model Deployment

Pretrained PyTorch neural network models will be deployed as smart contracts on both blockchains. These models will be trained on insurance datasets to predict claim risk and severity. The same models will be run in a traditional infrastructure to compare costs.

5.3 Cost Comparison

- **Blockchain Cost Analysis:** Gas fees for processing predictions on BNB Chain and Base Blockchain will be measured and recorded.
- **Traditional Cost Analysis:** Estimations of operational costs (including personnel, cloud computing, IT infrastructure, and software tools) for traditional insurance companies will be collected using publicly available data and academic reports.

5.4 Performance Evaluation

The accuracy, speed, and scalability of predictions on blockchain platforms will be compared to those of traditional systems. Metrics like mean squared error (MSE) for prediction accuracy and transaction completion times will be recorded.

5.5 Scalability and Security Analysis

Blockchain's ability to scale predictions to larger datasets and its security features will be analyzed and compared to traditional centralized insurance infrastructures.

6 Evaluation Strategy

6.1 1. Cost Efficiency Evaluation

The goal is to measure and compare the **cost per prediction** on blockchain platforms versus the estimated operational costs of traditional insurance companies for making similar predictions. This will include analyzing gas fees on the blockchain as well as calculating the infrastructure and personnel costs associated with traditional systems.

6.1.1 a. Gas Fees and Transaction Costs (Blockchain)

For each prediction processed on both BNB Chain and Base Blockchain, the **gas fees** incurred by running the smart contracts will be recorded. Gas usage will be measured in units of gas, and the corresponding transaction costs will be calculated by multiplying the gas used by the prevailing gas price in the network (in BNB for BNB Chain and ETH for Base Blockchain).

6.1.2 b. Operational Costs (Traditional Insurance)

Estimated **operational costs** for traditional insurance companies will be calculated using publicly available data and research reports. These costs include:

- **Cloud infrastructure costs**: Costs incurred by traditional insurance companies for using cloud services such as AWS, Google Cloud, or Microsoft Azure.
- **Personnel costs**: Estimated salaries for data scientists, actuaries, and IT staff responsible for building and maintaining predictive models.
- **Software costs**: Costs related to licenses for using predictive analytics software (e.g., SAS, SPSS), or cloud-based machine learning tools like AWS Sagemaker.

6.1.3 c. Cost Comparison Formula

To quantify cost efficiency, the **Cost Efficiency Ratio** will be calculated as:

$$\text{Cost Efficiency Ratio} = \frac{\text{Cost per Prediction (Traditional)}}{\text{Cost per Prediction (Blockchain)}} \tag{1}$$

Where:

- **Cost per Prediction (Traditional)** includes the sum of operational costs for predictions made in centralized systems.
- **Cost per Prediction (Blockchain)** refers to gas fees and transaction costs on the respective blockchain platforms.

6.1.4 d. Statistical Methods for Cost Analysis

- **Mean Cost:** Calculate the *mean gas fee* per prediction for multiple blockchain transactions to obtain an average cost per prediction on each blockchain platform.
- **Standard Deviation:** Calculate the *standard deviation* of gas fees across different transactions to evaluate cost variability.
- **Hypothesis Testing:** Use a *t-test* or *Wilcoxon signed-rank test* to statistically evaluate if cost differences between blockchain and traditional systems are significant.

6.1.5 Success Criteria

If the **Cost Efficiency Ratio** shows that blockchain-based predictions are significantly cheaper than traditional systems, this will support the hypothesis that blockchain is a more cost-effective solution for insurance predictions.

6.2 2. Prediction Accuracy Evaluation

The second key dimension is ensuring that the **prediction accuracy** from the blockchain-deployed smart contracts closely matches the accuracy of the original pretrained PyTorch models.

6.2.1 a. Accuracy Comparison

The **accuracy** will be calculated as the percentage of correct predictions out of the total, using the formula:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}} \times 100 \tag{2}$$

The **Accuracy Loss** will be calculated as:

$$\text{Accuracy Loss} = \frac{\text{Accuracy (PyTorch)} - \text{Accuracy (Blockchain)}}{\text{Accuracy (PyTorch)}} \times 100 \tag{3}$$

6.2.2 b. Mean Squared Error (MSE)

For continuous outputs, **MSE** will be calculated as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

Where y_i is the output from PyTorch, and $\hat{y}_i$ is the prediction from the blockchain model.

6.2.3 c. Statistical Methods for Accuracy Analysis

- **Paired t-test or Wilcoxon signed-rank test**: Used to evaluate if the difference in accuracy between PyTorch and blockchain models is statistically significant.
- **Threshold for Accuracy Loss:** A predefined threshold of 5% accuracy loss will be set.

6.2.4 d. Cross-Validation

Both PyTorch and blockchain models will undergo **k-fold cross-validation** to ensure the generalizability of accuracy results.

6.2.5 Success Criteria

- The **Accuracy Loss** between the PyTorch and blockchain models should be less than 5%.
- No significant differences in accuracy (as evaluated by MSE) should be observed between the models.

7 Expected Outcomes

- BNB Chain is expected to provide the most cost-effective solution due to its low transaction fees, making it ideal for real-time insurance claim predictions.
- Base Blockchain is anticipated to offer a more secure and Ethereum-compatible platform with reduced costs, suitable for regulated environments like insurance.
- Blockchain deployment is expected to result in a reduction of costs compared to traditional insurance companies, while maintaining accuracy and providing additional benefits like transparency and decentralization.