1. Introduction

Background

In today's world, the shortage of graphics cards has caused much concern and frustration for people who used computers as their primary tools for jobs. The high cost of these cards makes it difficult for people to afford them, hindering their ability to play games or create content. The fluctuating prices of GPUs further exacerbate the problem, and NVIDIA, one of the leading producers of these cards, must rely on third-party manufacturers for their chipsets. Manufacturers experiencing disruptions in their operations has led to a scarcity of graphics processing units (GPUs) and longer wait times for their production. As a result, the cost of GPUs such as the NVIDIA RTX 3090, which originally had a suggested price of $699, has skyrocketed to as much as $2,400 overnight. This scarcity and cost of graphics cards is a pressing issue that requires attention [1].

For those reasons people are trying to find the perfect time when they can buy a GPU. The forecasting method can be the way to solve the problem. Forecasting is a process of predicting based on historical data and extracting trends that can be approached using statistical or machine learning [2]. In [3]they study the GPU NVIDIA GTX 1060, which is affected by the bitcoin price. In this research, they are using linear regression models to forecast the upcoming price of GPUs. They found that the bitcoin's price affects the GPU's price. Another research that some researchers from CEEJ have done showed that the price of the GPU stock could be forecast using an optimal machine learning technique, the Nested Cross Validation algorithm [4].

One way to approach forecasting GPU prices is to use a deep learning model. This paper uses the recently developed transformer model initially developed to solve the NLP problem. In the transformer, from the input sequence, the model determines what other parts of the sequence are essential at each step [5]. The transformer has two parts: the encoder and the decoder. Theoretically, the transformer will use historical data to predict the upcoming prices in the experiment that some researchers have done. They are comparing the transformer and RNN. Transformers show up with excellent results and significant improvement [6]. In another experiment comparing transformers and LSTM, the transformer came out with a huge benefit because it is more stable and doesn't need so much time to train [7]. For this research, we will use the encoder layer to forecast the time series data that we have collected from keepa.

Problem Statement

Given the background information provided, this study will address the following issues. The problem statement for this project is split into two main components. The first part focuses on utilizing a transformer model to forecast time series data of GPU prices. The transformer model is a state-of-the-art machine learning technique that has been demonstrated to be successful for a wide variety of natural language processing and time series forecasting applications. This project aims to investigate the use of this method for predicting GPU prices. The second part of the problem statement centers on the prediction performance accuracy using a transformer model. This includes evaluating the model's capability to accurately forecast future GPU prices using historical data. The objective of this analysis is to determine the performance of the transformer model compared to other commonly used time series forecasting methods. The ultimate goal of this project is to develop a model that can accurately predict GPU prices, providing valuable insights for industry stakeholders.

Objective

In In this study, we are focusing on how a good transformer can predict the prices of GPU. Besides that, we also compare the model with other architectures, such as LSTM and RNN. The transformer itself is designed for forecasting sequential data. By using a transformer, the prediction can be done faster since the process in this model only runs once. Also, we limit the GPU variation so that there will be only 1 GPU that will be used for this research. Additionally, we only take the data from the first time the GPU is launched; September 2020, until November 2022. The result of the prediction will be quantified by using Coefficient Correlation (CC), Root Mean Square Error (RMSE), and Mean Averaged Percentage Errors (MAPE). All these metrics will evaluate the result of the model we have made.