Forecasting the price of AWS On-spot instances using Deep Neural Network Architectures

RANJITH KUMAR JAISHANKAR

Applied Research Project submitted in partial fulfilment of the requirements for the degree of M.Sc. Data Analytics at Dublin Business School

Supervisor: Dr Shahram Azizi

January 2022
Declaration

‘I declare that this Applied Research Project that I have submitted to Dublin Business School for the award of M.Sc. Data Analytics is the result of my own investigations, except where otherwise stated, where it is clearly acknowledged by references. Furthermore, this work has not been submitted for any other degree.’

Signed: Ranjith Kumar Jaishankar
Student Number: 10573397
Date: 10-01-2022
Acknowledgements

I would like to express my sincere gratitude to Professor Dr Shahram Azizi for his mentoring and constant supervision while I finished my master’s thesis work. I would also want to thank Professor David Williams for his assistance in determining a research topic in the first place. I would also want to thank my friends and family for their encouragement and valuable insights on the research.
Abstract

Among the cloud computing services, the concept of On-spot instance is the most popular which has been introduced by Amazon AWS, in order to utilize their spare capacity. On-spot instance follows the auction-based cloud model, where the price on-spot changes with time. In general analysis, it has been found that AWS On-spot instances are 30-40% cheaper than regular instances. The concept of dynamic pricing for AWS on-spot instance makes it complicated for some users, to bid for an optimal price. In order to help the users for selecting the optimal price with AWS On-spot instances, this research is predicted using the 4 different deep learning architectures which includes CNN, RNN, LSTM and Bi-LSTM for price prediction. To select the best performing model MSE, RMSE and MAE score has been calculated for each model over the test data. The better outcomes is achieved using Bi-LSTM model in terms of performance. In order to implement the concept of ON-spot price prediction, a web-portal using python flask has been developed which provides the predicted price of On-spot instance based on user input such as Region, Operating system, instance type, Time stamp etc.
# Table of Contents

Declaration ................................................................................................................................. 1  
Acknowledgements .................................................................................................................. 2  
Abstract ................................................................................................................................... 3  
Table of Contents ..................................................................................................................... 4  
List of Figures ........................................................................................................................... 5  

**Chapter 1** ............................................................................................................................ 6  
Introduction ............................................................................................................................... 6  
Research Question .................................................................................................................... 8  
Research Objective ................................................................................................................... 8  

**Chapter 2** ............................................................................................................................ 9  
Literature Review ...................................................................................................................... 9  
  Using Deep Learning Algorithms ........................................................................................... 9  
  Using ensemble learning algorithms .................................................................................... 11  
  Using regression algorithms ............................................................................................... 13  

**Chapter 3** ............................................................................................................................ 18  
Methodology ............................................................................................................................ 18  
  Data set description ........................................................................................................... 18  
  Data Pre-processing ........................................................................................................... 19  
  Exploratory Data Analysis ................................................................................................. 19  
  Feature Extraction ............................................................................................................... 22  

**Chapter 4** ............................................................................................................................ 23  
Model Training and Testing ..................................................................................................... 23  
Model Evaluation ..................................................................................................................... 23  
Design Specification .................................................................................................................. 23  
  Recurrent Neural Network Model ...................................................................................... 23  
  Convolutional Neural Network .......................................................................................... 24  
  Long Short-Term Memory Model (LSTM) ......................................................................... 25  
  Bidirectional Long Short Term Memory Model (BILSTM) .................................................. 26  

**Chapter 5** ............................................................................................................................ 27  
Implementation ......................................................................................................................... 27
Evaluation ........................................................................................................ 28
Evaluation based on Mean Squared Error (MSE) ........................................ 29
Evaluation based on Root Mean Squared Error (RMSE) .......................... 30
Evaluation based on Mean Absolute Error (MAE) ...................................... 30
Discussion .................................................................................................... 31

Chapter 6 ........................................................................................................ 32

Conclusion .................................................................................................... 32
Future Work ................................................................................................. 32
References ................................................................................................... 33

List of Figures

Figure 1 Methodology for On-spot Instance Price Prediction .................. 18
Figure 2 Frequency count of all regions ................................................... 19
Figure 3 OS Counts of all Regions .............................................................. 20
Figure 4 Instance Counts of all Instance type ......................................... 20
Figure 5 Comparison of Operating System based on Price .................... 21
Figure 6 Comparison of Instance Types based on Price ........................ 21
Figure 7 Comparison of Regions based on Prices .................................... 22
Figure 8 Architecture of RNN Model ....................................................... 24
Figure 9 Convolutional Neural Network Architecture ............................ 25
Figure 10 LSTM Model Architecture ........................................................ 25
Figure 11 BILSTM Model Architecture .................................................... 26
Figure 12 AWS On-SPOT instance prediction Web-UI ............................. 28
Figure 13 Predicted Price of AWS On-spot instance ............................... 28
Figure 14 Comparison of all models based on MSE Score ..................... 29
Figure 15 Comparison of all models based on RMSE Score .................. 30
Figure 16 Comparison of all models based on MAE Score ..................... 31
Chapter 1

Introduction

As innovation advances and develops step by step, there are new creations and redesigns that innovate the technological industry. Each course of the process and managerial activity gets easier to perform in a better way; then again, the data need to achieve this ease becomes huge. This makes a requirement for an immense and higher unit of the facility that has the capacity to create and house these huge loads of data with practically no issue or lags. Such data storing place in terms of hardware cannot be found in an easy manner and furthermore, whenever found, it comes with challenges and troubles alongside it. Such can be a ceaseless threat for the organization or association that intends to utilize it. For such complications, cloud seems to provide a very useful and effective alternative. With the cloud, the potential of going for higher storage, seamless working process, and hassle-freeness is possible. With the help of cloud, one can store heft amount of data for a very long period of time. Using cloud has also become very favourable for the users and big organisations as it offers the feature of paying just for the range of usage. This makes the cloud storage option seem lucrative and cost-effective.

When compared to the on premise data storing facility, the migration, storage and safety is very much optimised with the usage of cloud. Among the wide range of various sorts of cloud concepts, the most exceptional is Amazon's AWS. The memory, option of storage, and availability module it offers are no match to the other kind of storage options available in the market. The extremely recent and latest dispatch of AWS is the spot instance. The working of the spot slot is totally based on the techniques of bidding mechanism used by the buyer or the person who places such bids.

Although there are a lot of advantages rendered by the spot instances the purchasers, there is additionally a little intricacy that shows up with the techniques of bidding. The spot is given to the bidders who places a nearest value of the spot according to the estimates of it. When the bidder places the range of bid, which is very much higher than the price of the spot instance, then the Amazon spot is awarded to that particular user. But there is a major take in it. The dynamic nature of the spot instance plays a very important role in here. The price of the spot instance is not place on a fixed basis. Due to the changing trends and evolving practises of the market place and demand, the price of the spot instance differs drastically. Sometimes the price seems to be very low and sometimes very high. Although this trait makes the Amazon cloud
very popular in the market place among the users. Most companies go for AWS because of its dynamic nature. However, there are occasional price difference in the spot instance, it is very much better and lower when compared to the on-demand instances. The rate of the spot instance tends to increase corresponding to the demand on the user side. If there is no demand from the user end, then the rate of spot-instance is observed to be very much lower. So, the balance is maintained in that manner.

During numerous situations, such unanticipated changes in the value of the spot instances will more often occur after the offering of the bid by the user. If such happens, the spot instance will be completely closed by powerfully shutting down the action. The spot bidding protocol makes the client aware of progress, overhaul, or reinforcement of the data present on the spot during that specific occurrence. When examined and utilized the spot in a brilliant way the client can acquire higher benefit from it. However, the interruptions and rate differences in the bid can cause an incredible issue and complicated situation. The principal justification behind such variations and problems in spot instances is anticipated to be a direct result of the errors and deliberate blunders that occur in the allotment cycle.

Additionally, because of numerous higher price estimations and lower demand, numerous servers and spot occurrences in the cloud stay ideal for a long time. The cloud specialist and service providers find it hard to allot the right spot instance for the best necessities and prerequisites of the clients. This intricacy additionally prompts the issue of more problems in the allocation of spot instances to the user. From the customer's perspective, the spots that are accessible in the server will make the cycle easy of utility if any mishapening is situated out-of-bid place at whatever point at any continuous period by the spot instance. Spot reliability relies upon the market cost and the client's most extreme bid proposal rate. A very effective and efficient price prediction model is the need of the hour. When the prediction model is designed, all the left-out factors must be taken into consideration. A prudent price prediction model will clear out the problems and smoothen the process of bidding of the spot instance in an optimised manner. Using the normal standard time series model for the price prediction process will not be of any use because of the unequal time intervals of the slot and dynamic volatility of the spot instances. Since the price of the spot instance varies very dynamic multiple times throughout the day, the price prediction model must be very capable of predicting the price of spots one day ahead sufficiently and accurately.
At the point when the cost has precisely anticipated the organizations or the client can wisely stay away from the missing of spot occurrence because of offering in the inaccurate value range, which thus additionally helps the decrease of the expense of execution. In this paper, our sole main is to formulate and design an effective price prediction model of the AWS spot instance using deep learning algorithms. The plan is to increase the efficacy to predict and foretell the spot prices in tell processing time, with lesser errors and maximum efficiency. It additionally empowers the client to settle on their bidding price choice effectively. Many existing research of the past utilizes the strategies of advanced concepts like machine learning to anticipate the cost of the amazon spot instances. In this research, The 4 deep learning based neural network architectures is used for predicting the price of on-spot instances that are LSTM, Bi-LSTM, CNN and RNN. Along with it, a web application also has been developed using python and flask for forecasting the price of AWS on-spot instances.

Research Question

Based on our findings, this research will answer the following research questions.

1. Based on the user input such as Timestamp, operating system, instance type and region, what can be optimal price of an AWS On-spot instance?

2. Which algorithm can minimize the difference between actual and predicted price of AWS instance over the test data?

Research Objective

The main objective of this research is to help the users for bidding on the optimal price of AWS On-Spot instances. This research also aims for identifying the optimal deep neural network architecture by analysing the historical pricing data of AWS On-spot instances.
Chapter 2

Literature Review

This section of the research contains most of the existing research models and papers that are associated to the price prediction of spot instances. The research works that involves machine learning techniques, ensemble learning and other approaches are also analysed and discussed below.

Using Deep Learning Algorithms

Using neural networks are very popular for the prediction process. One such research done employing the neural network to foretell the price of the spot instance is carried by (Khwaja et al., 2015). The use of neural network and other deep learning algorithms contains making various protocols of data by testing subjectively with replacement, setting up a neural technique association on each data record, and averaging the results got from each pre-arranged neural method. It reduces the increasing errors which are differently related to using a singular neural network for the load. Models with certified data show the ampleness of the proposed procedures by displaying that using the strategies of neural organizations can decrease load checking methods, diverged from various existing techniques as shown by the research of (Veena et al., 2017).

The working of the spot instance is a long interaction and a top to the bottom arrangement is extremely fundamental and essential to foresee the highs and lows that occur with the spots. It is also equally essential to learn and know about the complete working process of the Amazon spot instance before proceeding with the prediction model. Although there are many research works stating the same, a paper done by (Ben-Yehuda et al., n.d.) gives a clear picture on what has to be given more preference and laid emphasis on. The cloud providers have tremendous measures of extra cut-off ranges to either help clients to get it or to suffer losses. Amazon is the essential cloud provider to address this test, by allowing clients to propose on as far as possible additionally, by surrendering resources for bid placers while their offers outperform an at times changing spot cost.

Amazon communicates the spot cost at this point doesn't divulge how it is settled as mentioned in the research by (Khan, 2017). By separating the spot esteem accounts of the cloud, the creators sort out how expenses are set and fabricate a model that produces costs solid with
existing worth follows. They inspect and observe that expenses are ordinarily not market-based and put as to a great extent as of late acknowledged. This planned model could assist clients with making wise bid offers, to the cloud providers plan advantageous structures, and experts design assessing computations. Thus, to make this work in the paper done by (Al-Roomi et al., 2013), they utilize the calculations of the method of deep learning is used in this study. From the investigation and the results that are totally founded on every one of the elements and outer properties, they show up to the referenced end.

Doing the cost optimisation is also very important when considering the amazon spot instance and the knack for it is given in the research done by (Lucas-Simarro et al., 2015). The authors performed an in-depth study by employed the advantage of neural networks and deep learning algorithms to find the best enhanced model to figure the exact cost range of the spot instances. In the paper, the authors state the amount of cloud spot providers in the total processing market and the demand of such spots is extending at a quick speed. They give out a wide extent of assessing plans, different kinds of cases, or even uncommon value-added features to differentiate from various competitors, making the cloud setting more many-sided and estimating the range of costs of the cloud instances. To understand the range of prices and the correlation of the demand, the paper of (Danysz et al., 2020) is analysed. In his work, the author portrays the dynamic nature and the drastic pricing difference of the amazon spot instances. At the point when the unexpected change takes in the spot instance case, the bid put by the client goes invalid for a while and makes it extremely difficult for the customer to get the specific spot occurrence they needed. To tackle this issue of inactivity the creator proposes different methodologies and steps in his exploration model.

To stay away from the occurrence of prompt end the creator utilizes the strategies of deep learning algorithms in the work by (Liu et al., 2020). He likewise defined an interesting sort of framework to do this methodology where Spot is taken as the main model. Neural networks and deep learning techniques are a portion of the exceptionally well-known strategies that are drilled by the examiners for a significant number of the expectation procedures and investigation in view of its precision and viability. One such paper that arrangements with the expectation of the cost of Amazon spot case is composed by (Agarwal et al., 2017) and it tells the idea of utilizing neural organizations for the equivalent. The expense of the spot or variable worth resources is the new progress in dispersed registering ideologies.
The spot price assessing instrument follows a closeout-based cloud system and formulated a model where the expense of spot cases develops with time and over a period. Here a maximum of 60 days is considered to be the time frame of the research. The cloud consists of different information and servers present alongside the spaces that are to be explored to give out the precise scope of forecast expected by the client. To make this thing a reality, a total outline is given by (Alkharif, n.d.). In this exploration, the author gives out the demonstrating and preparing process by total analysis of the cloud in a more appropriate and successful way. The appropriated registering resources are presented in various constructions, and a flood of handling resources is given at a more affordable expense. A super dispersed registering dealer, Amazon Web Services which is called AWS to put it plainly, gives such cunning resources as the spot event whose worth changes effectively subject to the resource interest from customers.

**Using ensemble learning algorithms**

To demonstrate the employment of price prediction process using the ensemble learning algorithms the research work done by (Yu et al., 2015) gives you a more elaborate understanding. Although the author has worked on the fuel price prediction process all other principal methods are the same. The technique for this value prediction process fills in as same as the value forecast of spot occasions. In the above-said data the method changes, all analysed models are totally examined to research the secret data characteristics and are in like way reproduced into some huge portions. For portrayal and affirmation, the previous information of the oil spot costs is used as the model data, and the specific results show that the proposed model quantifiably outmanoeuvres all recently referenced benchmark models including notable machine learning methods. The combined technique is also quite popular in the prediction process.

The research done by (Budgaga et al., 2016) using the combination of analytical, statistical as well as ensemble learning algorithms to foretell the price range of the spot instances. In the research work, the creator utilizes a troupe kind of ensemble learning strategy alongside some factual techniques and occasion reproductions tests to help the figured model. The work of prescient examination is likewise done in this exploration. A portion of the computational essentials regularly power limits on the extensiveness or possibly significance of investigation that can be driven with a discrete event re-authorization or a specific sort of re-enactment. The work portrays the technique for using the immense measure of figuring and limiting resources available in both private affiliations and public spots to enable continuous examination of discrete event diversions.
The research work of (Oliveira & Torgo, n.d.) gives a detailed analysis of using time series forecasting along with the ensemble learning algorithm for the prediction process. The paper portrays one more sort of ensemble process that objectives working on the judicious execution of these systems in time series forecasting. The machine learning calculations are seen as potentially the most ideal way of managing assumptions to do and play out the tasks. Past speculative studies of algorithms and patterns have shown that one of the basic clarifications behind this show is variable among others. A couple of methodologies exist to deliver assortment. The indispensable thought about the work which is acquainted here is to propose one more kind of assortment age that researches some specific properties of time plan assumption tasks.

The forecast of spot instances and their demand is a very long and monotonous interaction which includes the detailed examination and complete learning system about the AWS working cycle. One such review is finished by (Fernández-Cerero et al., 2020) that gives more understanding with regards to the work processes and the interaction that occurs in the cloud when the spot characters and elements change as per the streams and patterns that occur on the lookout. Waiting for the surge and group networks have turned into the establishment of enormous Cloud organizations, giving basically boundless flexible and versatile computational and limited resources. The journey for the capability and smoothing out of resources is one of the current key plots for gigantic specialist co-ops and is getting progressively tested since new handling principles of ensemble learning algorithms.

One more sort of comparative exploration that utilizes the integrity and advantage of ensemble learning calculations can be found in the paper introduced by (Ghani, 2005). The paper forwards the model and planned methodology at the forecast of the cost and protection of the closeout processes that happen online with all the due techniques that occur as an afterthought. Online sell-offs are making one more class of data about internet-based trades. This data fits a combination of employments and organizations that can be given to the two buyers and sellers in web-based business habitats. The data is assembled from online closeouts and couple of course of action computations is used to anticipate the conceivable end expenses of online deal things.

To track down the better course of the spot forecast, its accessibility, and its overall instrument in a compelling way the review did by (Jangjaimon & Tzeng, 2015) does a great deal of exploration. In this paper, the creator keeps forward a more viable and huger course of AWS
programming and working. Numerous viable procedures are additionally introduced here with respect to the method of bringing down the measure of the cost that is being utilized in the functional interaction. The proposed work presents the arrangement and execution of the further developed adaptable consistent check pointing for a multi-space application that confidently runs on the particular method under spot event esteeming.

The working of the ensemble learning methodology and its algorithms are clearly explained in the work done by (Alhamid, 2021). Here the author talks about the ensemble models and clarify the underpinning working of it. The detailed explanation sheds lights on the numerous areas of the technique that corresponds to better understanding of the topic. He also performs a comparison stating the working of this method on various technique and applications. The boosting, stacking and blending are also discussed in the work in an enhanced manner and the prediction process is also kept forth. This therefore states the effectiveness of ensemble learning in the prediction process.

Using regression algorithms

The regression algorithms are another very important aspect of technique that is being used by other fellow researchers to predict the price and dynamic nature of the spot instances. The range of errors and false predictions are much lowered when it comes to employing the regression algorithms which is clearly stated in the research done by (Mishra & Yadav, 2017). For analysis the authors take the comparison of the Amazon spot instances. The work depicts a similar idea of anticipating the cost of spot example utilizing relapse and regression techniques for an alternate reason. Typically foreseeing the cost of a spot occasion at a specific time frame is confounded. The matter with the investigation is that it ought to be done consistently without achieving any break in the organization or cycle. In the examination paper clarified, the creators investigate the expense of spot occurrences and the cycle is separated into basic stages and a backslide computation is used to expect significant information for cloud clients and cloud traders who need to start to benefit from the utilization of spots. One more significant technique utilized here is the regression section of calculations is presented here. It makes the estimation simpler and more powerful.

An exploration paper introduced by (Su et al., 2019) gives greater lucidity on the utilization of such techniques for the cycle of prediction. Here the creators define a functioning model that utilizes the Gaussian regression model at the expectation of gas cost in the most unpredictable and constant environment. They have likewise utilized a portion of the natural machine
learning procedure for the better working of the planned model. Such procedures are support vector machines, slope helping, and neural organizations. They secure substantial information and afterward happen with the preparation interaction. Yet, the strategy for regression will in general show all the better exhibition and exactness in the yields when contrasted with the other techniques of machine learning.

Another in-depth research carried out by (Shakhla et al., n.d.) gives a technique of using linear regression to predict the prices of the stocks. Here, only the scenario and analysis are different and the method is completely the same. The historical backdrop of the spot case is likewise remembered for this examination. The numerical estimation and investigations are done based on k nearest neighbours, one of the exemplary regression models that are best for the prescient examination kind. The consequences of this model are then traditionally contrasted and different portrayals of their investigation like vector machines, time-series forecast model, and the linear regression technique. The precision is seen higher to be in the relapse models. The linear regression strategy model is additionally perhaps the best strategy that can be utilized at the near cost forecasts study done here.

Another work that gave bargains on the stock expectation component, where the procedures of regression technique is utilized by (Sahoo, 2015). These days stock expenses are affected in light of various reasons like association related news, political, social-proficient conditions, and disastrous occasions very much like the spot occurrence which is taken for investigation. A lot of examination and studies were performed for the estimate of stock record regards similarly as the regular heading of progress in the rundown. Such endless models were made for anticipating the future expense of stocks anyway everybody has its own defects. Advanced insightful methodology going from just mathematical models and expert structures to neural associations have moreover been used by money related foundations. In this work, the creators inspect to predict the stock expenses using an autoregressive model. This sort of regression model is used because of its ease, dependability, and value.

One such comparative examination work introduced by (Moreira et al., 2012) clarifies the working of regression procedure for a wide range of prescient and ensemble methodologies. The goal of the model is to work out a backslide of the regression models and to merge a couple of models to work on the estimate accuracy in learning issues with a numerical objective variable as a successful objective. The pattern of the learning system can be divided into three phases of working: the inception stage, the programming stage, and the blend stage. The
creators talk about different ways of managing all of these stages that can deal with the backslide issue, orchestrating them in regards to their appropriate characteristics and associating them to responsibilities from different fields. In addition, this work makes it possible to recognize interesting areas for future investigation.

One such examination work given by (Xiong et al., 2014) tells about the procedure of span estimating process utilizing relapse technique and regression to draft the model structure. In here, the creators utilize the famous help vector machine to bivariate the examined tests alongside the issue of the expanding interest for the accessibility of power. This proposed exhibiting framework is improved with the advanced methods to separate simultaneously both the cut-off points time course of action, created in sorts of complex-regarded time game plan, of force interest on a month to a month out of every hour premise, achieving getting the normal relationship among lower and maximum cut-off points. One such comparative examination work by (Larivière & Van den Poel, 2005) presents us the clearness of the functioning system of working course of relapse calculations and random forests methods for the expectation of maintenance. This limit is vital and essential for practically all organizations.

The assessment did by the creators makes huge discoveries and shows that both discretionary methodologies give better fit to the evaluation and endorsement test appeared differently in relation to normal relapse models of direct and calculated system models as per (Ravikumar & Saraf, 2020). In addition, through the exploration, the creators likewise found evidence that a comparative game plan of variables contrastingly influence buying as opposed to slipping away versus advantage lead. A portion of the discoveries and revelations suggest that past customer lead is more crucial for produce go over purchasing and great advantage headways, while the agent's occupation significantly influences the customers' tendency. In one of the extremely conspicuous and significant work introduced by (Criminisi et al., 2013)the creator gives out the total system and successful working course of utilizing relapse calculations and regression method. The creator takes up the undertaking of perceiving and distinguishing the constructions present in the CT examines utilizing the multi modular examination of the three-dimensional figures. Contrasted with the utilization of arbitrary strategy the paper referenced here utilizes the relapse technique which is characterized to be at the highest point of the characterization calculations. The affirms of the dataset and the preparation modules are additionally assessed and introduced in the correct way. As per its size and characterized aspects, the planning is additionally done deferentially.
The course of quantitative examination is likewise done and the outcomes got from the test tells the precision and adequacy of the relapse calculations. One extremely conspicuous exploration figures and insight given by (Guo et al., 2015) predicts the necessities and models that are needed to put the bid of the spot case in the correct way during more interest and much pinnacle hours. In the exploration the creators utilize the most fundamental and the fundamental AI procedures to manufacture the said model. The accessibility of the spot examples is obviously clarified in this paper by the creator.

The creators formalize and deal with this issue and propose an availability and cost careful framework which watches out on all offering instrument portrayed by (Dawoud et al., 2011). Preliminary outcomes show that their proposed component can diminish the costs of an appropriated organization and a spread accumulating independently while at this point keeping availability level in equilibrium to it is by using on-demand events. From this exploration the instrument of offering and spot occasion accessibility is plainly settled.

One such comparable exploration paper is composed, tested and afterward introduced by the creator (Antipov & Pokryshevskaya, 2012). This paper manages the precise expectation component of accessibility of the space of private bequest and the examinations and increment or abatement drifts that occur with it. The creator widely utilized the procedure of arbitrary woodland to anticipate the working of this model. The course of experimental review is additionally done in this examination work. Notwithstanding every one of these, the regression investigation which has a similar methodology of utilizing decision trees and neural organization associations are additionally utilized on a higher scale. The adequacy of utilizing the techniques of machine learning calculations is very grounded in many investigates works and examination.

One such successful instrument of examination is available in the paper done by (Booth et al., 2014). In this element, the creator presents the model of decision trees to test and foresee the working of exact investigation, normalities and irregularity consequences for the vital information tests that are utilized for the monetary estimations. To make this exploration successful the creator likewise utilizes different writing works and conservative and monetary experiences. An expert system that uses the main methodologies of machine learning to expect the worth return over a portion of a periodic events and a short time later uses these gauges to develop a useful trading technique. One such paper that presents this technique of regression
calculations along with the method of random forest for the compelling forecast of offer market stock costs and gauges is given by (Khaidem et al., 2016).

Anticipating designs in monetary trade costs has been a space of interest for researchers for quite a while due to its complex and dynamic person. Trademark in the monetary trade of stocks overall makes the task of assumption testing. Expecting and showing the pace of its instability, though reasonable can't be the answer for the diverse extent of issues experienced in assumption, present second or something different.

The dynamic nature of market patterns, unequivocally connected with assessing goofs, ought to be restricted to ensure the insignificant risk in adventure as shown in the work by (Cai et al., 2017). The makers propose to restrict deciding screw up by viewing the issue of estimating as a gathering issue, a popular set-up of computations in Machine learning. So, instead of standard approaches, this model seems to be much better. However, the utilization of the techniques of regression plays an exceptionally critical undertaking and job in the expense enhancement process.

One such exploration work is finished by (Grzegorowski et al., 2021). The creator completes his examination and investigation process in the enormous information responsibilities to perceive how viable the method involved with displaying can be. The expansion of the cloud model is done according to the opening. The continuous investigation of the organization is additionally one without prior warning cycles. Intelligent data taking care of has accepted unmistakable quality as the establishment of the current business accomplishment, and it is worked with by Big Data arranges that give significance all things considered, limitless flexibility.

One such exploration done by (Mishra et al., 2019) shows the working of virtual machine calculations that have the division of spot case, on-request and saved examples. The measure of cost that is spend on these tasks are viewed as huge. In this examination work, an estimation has been expected for expecting the spot cost to work with the customers in offering the most ideal model. Added to this, there is likewise a pointing computation, which has been given by the creator for saving the task's progression at ideal time extends by the usage of the proposed spot esteem assumption estimation. The proposed estimations that are planned in this examination work, highlight the usage of handled data at the figure of expenses in short stretches. The outcomes of this research also look in a promising manner.
Chapter 3

Methodology

There are lot of variation in the costing of an AWS on-spot instance, depending mostly on the instance type, operating system, time, region, marketplace etc. Hence, prediction of accurate cost of the AWS on-spot price instance is the main aim of this research, using deep learning algorithms. There are a set of steps such as Data collection, data cleaning/Pre-processing, Data visualization, Feature Engineering, Model training and testing which are needed to be followed to generate the best possible outcome. The detailed discussion of each of these steps follows. The flowchart diagram in fig.1 represents the proposed framework for the process of prediction of the price of the AWS on-spot instance. The data of AWS on-spot price of 5 different zones is collected for this work and above-mentioned processes has been followed.

Figure 1 Methodology for On-spot Instance Price Prediction

Data set description

The dataset for this task has been collected from Kaggle AWS Spot Pricing Market (AWS Spot Pricing Market, 2022). The dataset mostly contains the data of 3 months of AWS on-spot price for each region. There is total 5 features in the dataset, namely price, Date Time, instance type,
OS, and information of the region. The total size of the dataset is 1.69GB, for 9 regions. For our work, only 5 regions are considered, due to limited resource capabilities. The data operations which include data pre-processing, feature engineering etc are performed on Jupyter notebook.

**Data Pre-processing**

The process of converting raw data into an understandable format in called data pre-processing. It is an important step in the analysis as no work can be done with raw data. The quality of the data needs to be checked before implementing deep learning algorithms. 20,000 rows is collected from each region and concatenated the data for all the five regions. Our Data pre-processing step includes data cleansing, normalization, eradication of null values etc. Categorical features like instance type, instance region and operating system were efficiently handled with label encoding techniques. A detailed description of each feature is described in succeeding sections.

**Exploratory Data Analysis**

After pre-processing of dataset output data contains 100000 samples with 9 features. In order to extract out important and meaningful information from data, visualization and data analysis is executed on the data. In the first analysis frequency of data based on regions is analysed. Where, It is observed that in terms of region EU-central-1a is highly used while sa-east-1b is used at least as shown in figure 2.

![AWS region wise price collection](image.png)

*Figure 2 Frequency count of all regions*
On analysing the operating system for spot-instances, it has been observed that 45.9% users use Linux/Unix while 33% users use SUSE Linux and remained 21.1% users appreciate Windows as their operating system on AWS Instances as shown in figure 3.

![AWS OS Counts](image)

*Figure 3 OS Counts of all Regions*

During analysis of the whole data, instance type used by users, it is observed that instances m4.2xlarge and m3.xlarge are highly used and instance type such as p2.8xlarge, p2.16xlarge are least used instance types. The Frequency graph for instance type is shown in Figure 4.

![AWS Instance-type](image)

*Figure 4 Instance Counts of all instances type*
During analysis based on the Price and operating system. It is observed that price of windows is more in comparison to SUSE Linux and Linux/UNIX. Linux 11 operating system is available at cheapest rate in for AWS On-spot instance. The graph is shown in figure 5.

![Figure 5 Comparison of Operating System based on Price](image)

In next analysis based on price and instance type and regions it has been observed that for instance type the instance x1.16xlarge, hi1.4xlarge, x1.32xlarge, i3.16xlarge, p2.8xlarge and p2.16xlarge type of instance has more prices in comparison of all other available instance as shown in figure 6 while based on regions ap-south-1a price is highest while sa-east-1b has lowest price as shown in figure 7.

![Figure 6 Comparison of Instance Types based on Price](image)
Feature Extraction

For achieving more accurate results, feature extraction, which is an important step, is performed. The extraction of features like month, week, day, hour, minute, second etc. is done based on timestamp feature. Other features such as region, sub-region, instance type, and operating system are considered for analysis. According to our observations four features were categorical in nature which then converted into numerical values using label encoder. After this step, mapping of different numeric values to different categorical values in the dataset is done.

Figure 7 Comparison of Regions based on Prices
Chapter 4

Model Training and Testing

For this work four different deep learning algorithms are implemented in this project for the analysis and comparison, and dataset is split into training and test dataset in the ratio 80:20 where training dataset is used to train each implemented algorithm and test dataset is used to test the performance of algorithms. The four deep learning algorithms are simple Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Bidirectional LSTM models and each model is trained over the dataset using Jupyter notebook. The epoch size for model training is considered as 10.

Model Evaluation

After training of each model every model is evaluated with other models based on different metrics. This step is important because this step enable us to identify the best performing model. The best model can be used in order to predict the on-spot AWS instance price. Since it is a regression problem therefore the Mean Absolute error (MAE), Mean Squared Error (MSE) and Root mean squared error (RMSE) are used to evaluate all implemented models. Lower values of all these metrics means better model performance. After complete training model is tested on test data and test MSE is calculate which is used to evaluate each model. For further analysis and better understanding, the MSE, RMSE, MAE of all models is plotted in bar plot to make comparisons.

Design Specification

For this work four different models are implemented for prediction and analysis therefore the architecture of all models is different. Simple RNN, CNN, LSTM and BILSTM Model is used. In further subsections all the models are described briefly.

Recurrent Neural Network Model

It is a class of neural network which is good at modelling time series data i.e., sequential data. for loop is used by RNN layers to iterate over of a sequences or time steps, encoded information about time steps seen so far while maintaining an internal state. Network is first provided with a single time step. After this calculated state along with current state a new state is calculated, and this continues till forward propagation of network. The output is calculated when all time steps are completed and then this output is compared to the actual output and an error is
generated which is then back propagated to the network to update weights and in this way RNN is trained. The architecture as shown in figure 8.

![Architecture of RNN Model](image)

**Figure 8 Architecture of RNN Model**

**Convolutional Neural Network**

A convolutional neural network does the work of the visual cortex for computers. Just like a human brain takes input from whatever the eyes see and processes it to classify and differentiate from other images CNN takes input in the form of images breaks it in the form of a matrix, add biases to each section of the matrix in order to classify it from different images and then gives output as the result. CNN is used for image classification and recognition. CNN tries to mark and remember features of an image and later compare those features with the features of another image to differentiate between the two. The part with more prominent features will have more bias added to it and the sections that are empty or have fewer features will have reduced bias as it is not accounting for the main image as much as the sections having more features. This helps the CNN model in classifying images.
Long Short-Term Memory Model (LSTM)

LSTMs are designed to solve the problem of RNNs that is Recurrent Neural Networks. Now, RNNs have short term memory so if you are using RNN for translation from one language to another or you are using it for autocomplete algorithms it may miss out on important information due to its short-term memory. LSTMs solves the problem of RNNs as it knows which information to remember and which to forget. LSTMs consists of gates, if the algorithm thinks the information is useful the gates let it pass to the memory and if it thinks the information or the word is not accounting to the end outcome it will ignore it. In this way, LSTMs are able to remember useful information. LSTMs are also fast and consume less computational power as compared to RNNs.
Bidirectional Long Short Term Memory Model (BILSTM)

BILSTMs are similar to LSTMs in such a way that they both save specific information to the memory and ignore the rest of the information. BILSTM algorithms have more context about the dataset as compared to LSTM because here information is fed from both ends, from start to end as well as from end to the beginning that gives more information to the algorithm for processing and predicting the right outcome. Now, as the BILSTM is using 2 models for training one from the beginning and one from the end, there must be a way to combine these two for processing the outcome. In order to merge BILSTM uses 4 techniques that are addition, multiplication, concatenation and averaging. It uses these methods to combine the two algorithms and produce better results.

Figure 11 BILSTM Model Architecture
Chapter 5

Implementation

In this project a web-framework based on python -flask is used to produce a web - interactive user interface which is easy to use, interactive and transparent in nature. A user-friendly environment is developed in order to predict the price of AWS spot instances where a user will provide the information such as Week, Month, Day, Hour, Second, Instance type, Region, and required Operating system (OS). Developed Web-UI can be run by executing a piece of python code. Since four deep learning models (LSTM, BILSTM, Simple RNN, CNN) have been implemented and the best optimal model is chosen to make predictions in real time. In order to train and execute each model various python libraries have been used. These libraries consist of NumPy, pandas, matplotlib, Sklearn, TensorFlow, Keras, Flask etc. Each deep learning model is initialised and trained by the use of Keras and TensorFlow while NumPy, Pandas, Matplotlib and seaborn libraries have been used to do data pre-processing, visualization of graphs and exploratory data analysis. The best optimal model is chosen based on evaluation of models on different metrics. Predictions are performed based on the best model weights which are saved into a. hdf5 file. In order to generate the interactive user-based web interface some other languages and technologies are also used such as JavaScript, CSS, and HTML. The developed application can be run on port number 5050 on any system and can be deployed on any cloud platform for easy and long-term usages. Figure 12 demonstrates the snapshot of the executed Web UI. Figure 13 represents the predicted price for provided input data.

The system with following specification is required to run the application.

- Operating System: windows 10
- Random Access Memory (RAM): 12GB
- Number of CPU cores: 4
- Hard disk: 1 Tb
- Languages: HTML, CSS, Python
- Platform: Jupyter Notebook

After providing correct inputs to the system, the model will predict the minimum price for an AWS spot instance as output for that specific region.
Evaluation

In order to obtain the optimal model for prediction of AWS spot instances, an evaluation step is necessary; it will help us to select the best model for predictive analysis. Different deep
Learning models is executed in this project and since it is a regression problem therefore each implemented model is evaluated based on metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). First the whole dataset is split in the training and test in the ratio 80:20 and each model are trained on training dataset and hence evaluated on test dataset. Different deep learning models that have been implemented as follows: Simple RNN, CNN, LSTM and BILSTM. Every implemented model is trained for 10 epochs, tested on a test dataset, and compared on different metrics in order to select the best model for prediction of on spot instance price of AWS. Each evaluation is visualised using bar plots for better understanding.

**Evaluation based on Mean Squared Error (MSE)**

The mean squared error is used to measure the average difference between predicted values and true values of a model by squaring the error difference. MSE is considered as a risk function because it corresponds to the predicted value of the squared error loss. Values obtained are always non-negative and generally values are close to zeros. The MSE is known as the second moment of the error because it incorporates both the variance of the model and its bias. First model that is implemented is Simple RNN model when evaluated on test data shown MSE value 3.22 and second model implemented is CNN which shown MSE value 3.80, While on evaluating third model i.e., LSTM MSE value observed is 2.68 and for fourth model i.e., BILSTM MSE value observed is 2.14. The lowest value of MSE is obtained by BILSTM. The figure 14 shows the comparison of models based on MSE score.

![Figure 14 Comparison of all models based on MSE Score](image-url)
Evaluation based on Root Mean Squared Error (RMSE)

Root Mean Square Error (RMSE) is considered as a standard way to evaluate regression problem-based models. It is just like MSE, but it has values which are square rooted of MSE therefore if MSE produces large errors which are difficult to understand then RMSE acts as a better option for evaluation. The RMSE score obtained over test data by the Simple RNN model is 1.80 and CNN model is 1.94. For the LSTM model, the RMSE score obtained over test data is 1.63 and 1.46 for the BILSTM model. The lowest RMSE score also has been obtained using the BILSTM model. The figure 15 shows the comparison of models based on RMSE score.

![Comparison of all models based on RMSE Score](image)

Evaluation based on Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is used to evaluate models which are used for regression problems. It calculates the average difference between the predicted values and actual values. This is a scale based dependent accuracy and sensitive to outliers because it calculates error in observations taken on the same scale. The MAE score obtained over test data by the Simple RNN model is 0.88 and CNN model is 1.08. For the LSTM model, the MAE score obtained over test data is 0.61 and 0.32 for the BILSTM model. The lowest MAE score also has been
obtained using the BILSTM model. The figure 16 shows the comparison of models based on MAE score.

![Figure 16 Comparison of all models based on MAE Score](image)

**Discussion**

After executing four experiments above, it is concluded that the BILSTM model is a best optimal model in order to predict AWS spot instance price because it has provided the minimum MAE, MSE and RMSE outputs. An interactive user-based Web UI is implemented in which a user has to provide following inputs i.e. Second, Hour, Day, Month, Year, OS type, Region and instance type and based on these inputs the developed system will make lowest price prediction on instance spot of AWS for that specific region. It has been observed that for model BILSTM, values of MSE start to get constant values while for other models it is observed that after a certain number of epochs the Mean square error was found to be start increasing which shows the condition of overfitting. In this project work the dataset used in this analysis is balanced and therefore it has prevented the results from being biased towards the one kind of values/features.
Chapter 6

Conclusion

Anticipating price of spot instance is a task of the essence and still an unconquered domain of research. The horizon in which the concept of spot instance is very much rising and falling and is no more than derived on the basis of user preference on the underlying terms of the market and its demand. Still, there are a great deal of elements which contributes to the great extent of anticipating the price of spot instance, which is composed of the date-time, region, operating system, and instance type. In this research, all these important factors are considered in order to anticipate the price of AWS spot instance by providing a user an ergonomic environment using WebUI. The developed application can be very functional and advantageous for the customers, who wants to run the short-term job/application over a cloud spot instance. The developed application will also be of service to the customers to make a bid for the optimal price of spot instance in order to make use of the cloud services without engaging in any error and interruption. After evaluating all the experiment, it can also be concluded that BILSTM achieves the better results as compared to Simple RNN, CNN and LSTM model and calculate lowest MSE score of 2.14. Due to limited computing capacity of the resources and large training time of deep learning models, this experiment has been performed over the 100000 number of samples. However, in the future work the size of data can be increased and results can be optimized by hyper parameter tuning of models.

Future Work

Also, as the data of AWS on-spot price is always increasing in nature. In future work, this problem can be considered as a big data problem which can be resolved using distributed solution. Hadoop and spark are some of the prominent scalable solutions which can process the large amount of data in minimum amount of time by using distributed technology, where task is divided among the multiple workers to reduce the processing time.
References


