PAYMENT FORECASTING OF ETISALAT CUSTOMER IN YEAR 2022

TECHNOLOGY ACADEMY

SUBMITTED BY: Ahamed Siraj Lebbe

REVIEWED BY : Mohammed Marwan

Contents

1. Objective

The objective of this project to apply the machine learning techniques to a real-world problem which is to build a time series sequence model for forecasting the consumer customer's payments in year 2022. The forecast model is built based on last 10 years of payment data taken from the payment system of Etisalat.

The outcome of the payment forecasting will be helpful for the IT operation teams to plan and optimize the back-end resources in the IT systems in order to accommodate the extensive incoming traffic of payments forecasted for year 2022.

2. Introduction

Etisalat is the first and the largest mobile telecom operator in the United Arab Emirates (UAE) found in 1976 as a joint-stock company between International Aeradio Limited, a British Company, and local partners. In 1983 the ownership structure was changed to the government of United Arab Emirates, where the government holds a 60% share in the company and the remaining 40% being publicly traded under Abu Dhabi Securities Exchange (ADX).

Etisalat is offering telecom services to different customer segments in UAE market such as consumer and enterprise customers. The telecom giant is equipped with modern and innovative direct and indirect payment channels to serve its customers. Etisalat's higher revenue is gained through its customer payments towards services offered by Etisalat and it is well equipped with high performance of IT systems for post processing the payments upon received from the customers.

2.1 Machine Learning

Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make

classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics.

Machine learning classifiers fall into two primary categories:

- **Supervised Learning**

  It is also known as supervised machine learning, is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately. This occurs as part of the cross-validation process to ensure that the model avoids overfitting or underfitting. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. Some methods used in supervised learning include neural networks, naïve bayes, linear regression, logistic regression, random forest, support vector machine (SVM), and more.

- **Unsupervised Learning**

  It is also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, image and pattern recognition. It’s also used to reduce the number of features in a model through the process of dimensionality reduction; principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, probabilistic clustering methods, and more.

2.2 Time Series Sequential Data

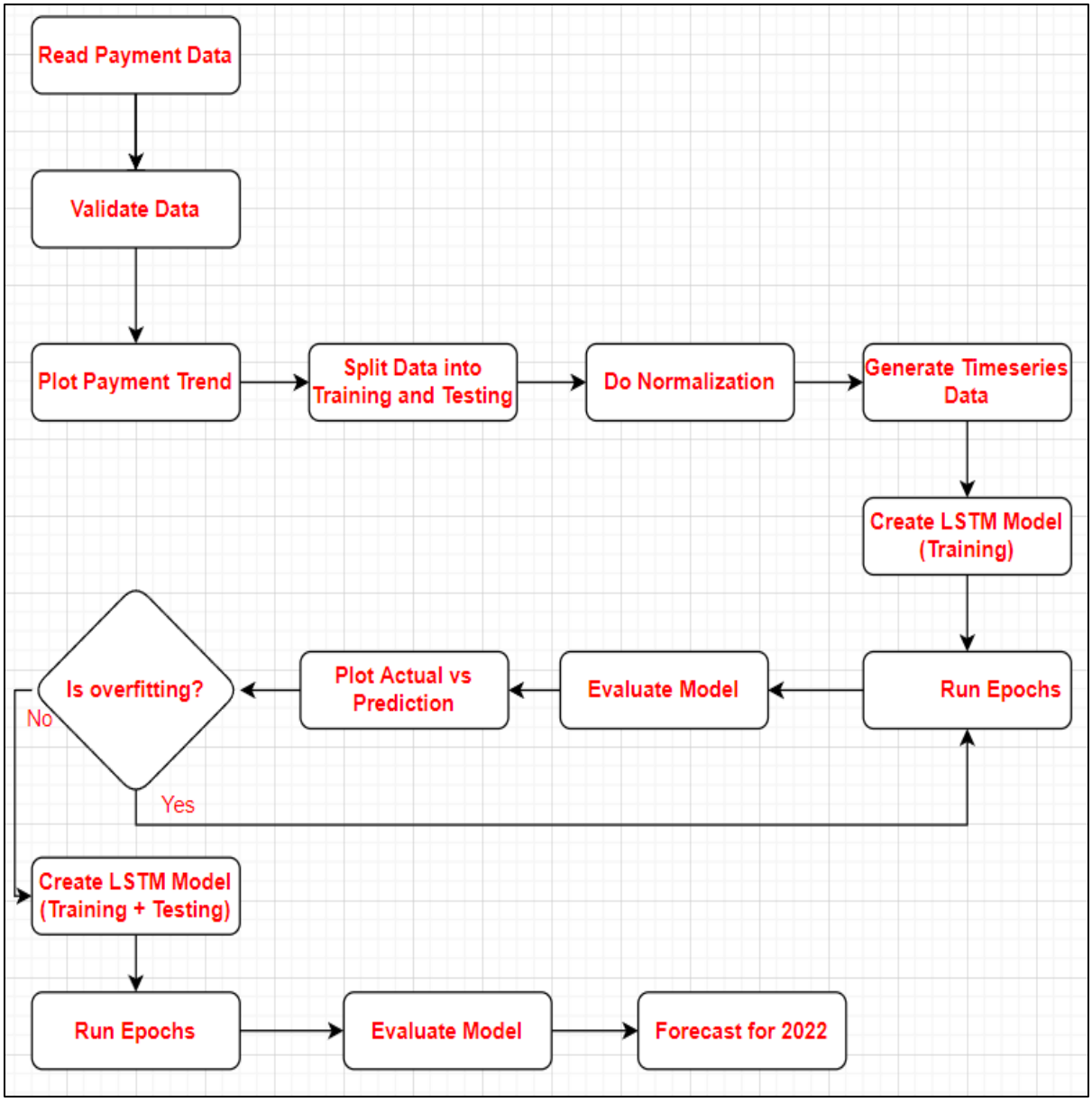
Time series forecasting use cases are certainly the most common time series use cases, as they can be found in all types of industries and in various contexts. Whether it is forecasting future sales to optimize inventory, predicting energy consumption to adapt production levels, or estimating the number of airline passengers to ensure high-quality services, time is a key variable. And, yet, dealing with time series can be challenging. The data, which consists of sequences of observations recorded at regular time intervals, may contain noise, be highly lumpy, or even be intermittent depending on the context

Unlike regression predictive modeling, time series also adds the complexity of a sequence dependence among the input variables. A powerful type of neural network designed to handle sequence dependence is called recurrent neural networks. The Long Short-Term Memory network or LSTM network is a type of recurrent neural network used in deep learning because very large architectures can be successfully trained.

3. Architecture / Workflow

In this project scope, a time series model “LSTM” network of recurrent neural network has been used to train the 10 years of payment data. The workflow of creating the forecasting model is depicted below.

Figure 1: Workflow Diagram of Payment Prediction

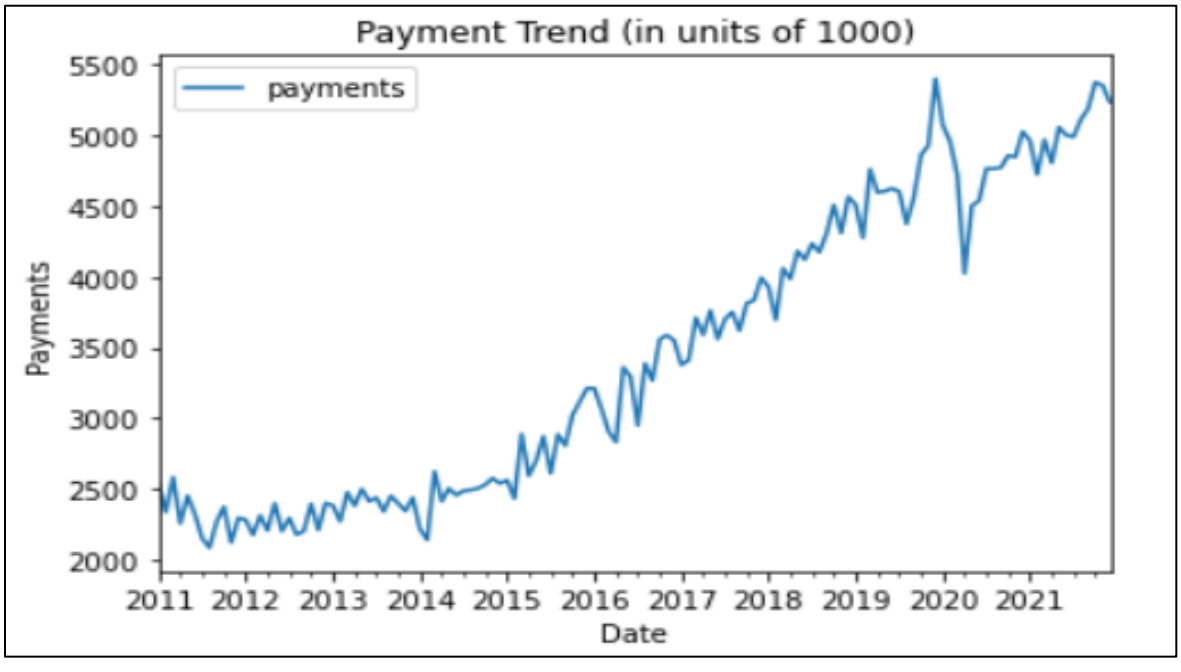


Source: Author

4. Results and Outcome

The payment trend of consumer category in the last 10 years is shown below.

Figure 2: Payment Trend in Last 10 Years (in units of 1000)



Source: Payment System

It is observed that the payment trend increases over period of time but there is a sudden decline in year 2020. This may be due to pandemic situation in UAE. The total payments in 2011 was approximately 27M where it was increased to approximately 60M in 2021 (more than 50% increase in 10 years' time). Based on this data, the average monthly and daily payments in 2021 are given below.

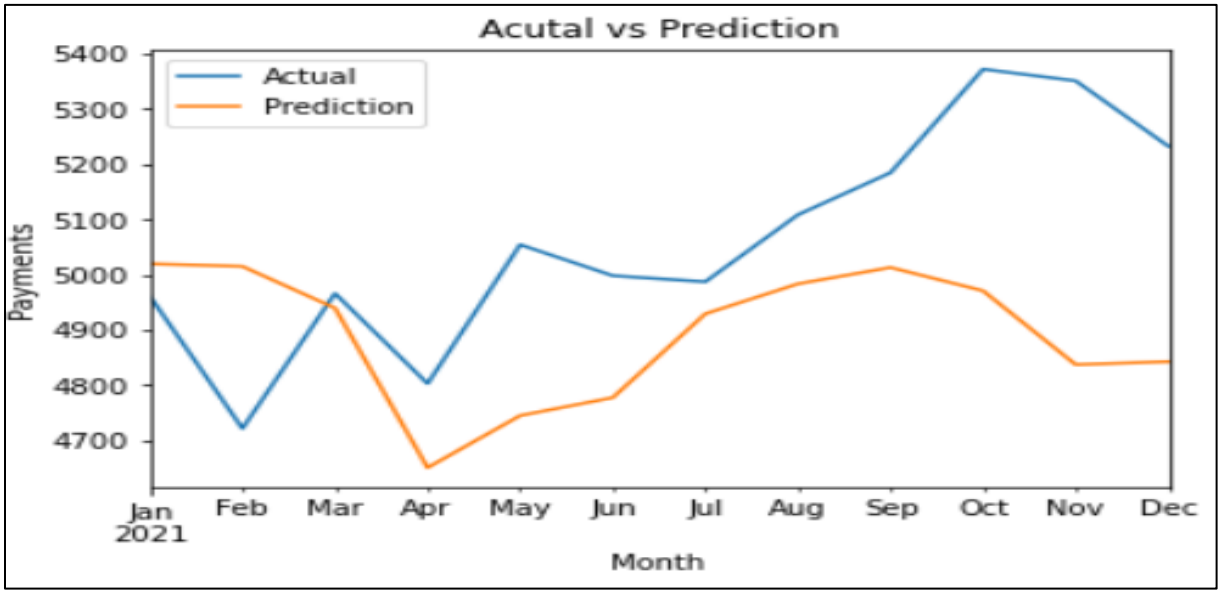
Table 1: Payment Inflow in 2021 (Actual)

Description	Average Payments (Counts)
Total payments in 2021	60,000,0000
Monthly payments	5,000,000
Daily payments	170,000

Source: Payment System

After successful trained the 9 years data till year 2020, the LSTN model was used to predict the payments for 2021 so that the accuracy of the predicted model can be validated with the actual payment data of year 2021. The below plot shows the trend of actual vs predicted for year 2021.

Figure 3: Payment Trend Actual vs Predicted for Year 2021



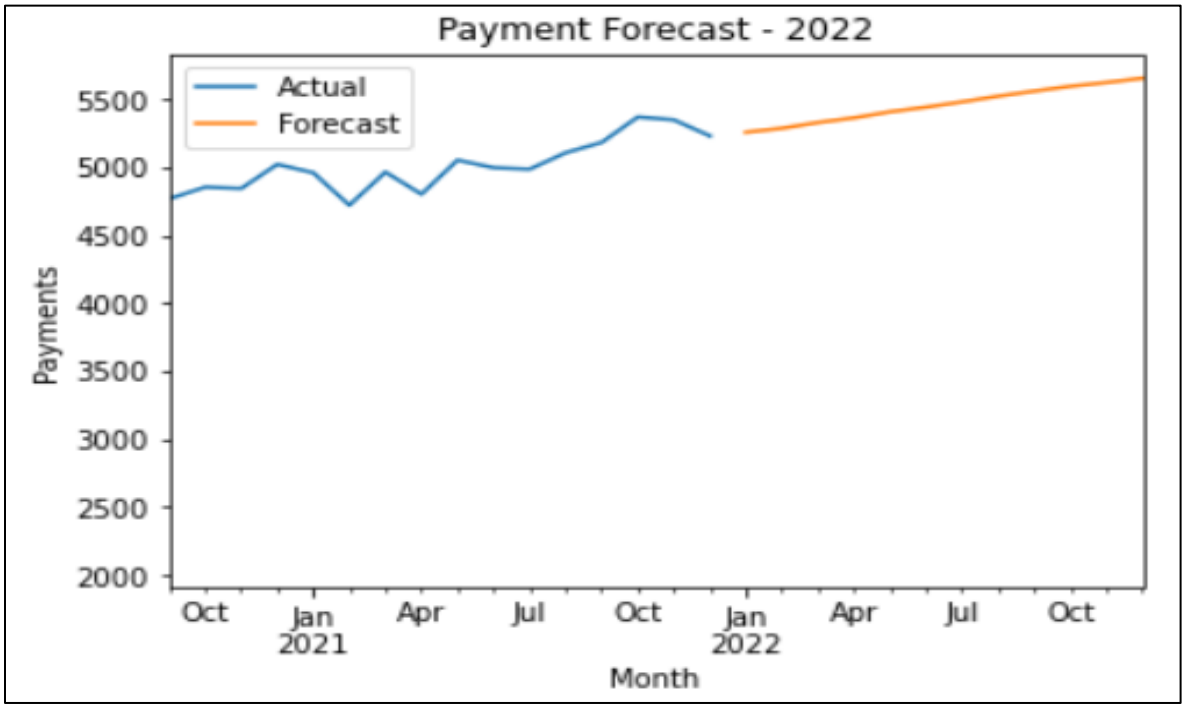
Source: Author

The above plot graph shows a small variance in the payment trend over actual vs predicted but it is not much outfitted. Therefore, the LSTM model was again created merging the training data (till 2020) and the testing data (2021) then the payments was predicted for year 2022. The below plot graph shows the predicted payments for year 2022.

Figure 4: Payment Prediction for Year 2022

Source: Author

Figure 5: Payment Forecast for Year 2022



Source: Author

The prediction shows that approximately 66M payments is expected in year 2022 which is approximately 10% increase in the consumer payments compared to 2021. The predicted additional payment traffic is summarized as below:

Table 2: Additional Payment Inflow in 2022 (Predicted)

Description	Average Payments (Counts)
Additional Monthly payments	5,000,000
Additional Daily payments	17,000

Source: Author

5. Future Work

This project can be extended to the payments of enterprise customers and the payment forecasting can be determined for year 2022.

6. Conclusion

This report analyzes the current payment trend of consumer customer payments and predicts the payments in 2022. The last 10 years data considered for this report which was taken from the payment system of Etisalat. As of 2021, the payment system is equipped with below IT infrastructure.

Table 3: IT Infrastructure Details of Payment System

Weblogic Servers	No of servers = 4	No of clusters/instances = 16 (=4*4)
Oracle Database Server	Physical Memory = 273 GB with 40% of current utilization	

Source: IT Operation Team

The incoming payments are distributed equally among 16 Weblogic instances on robin round basis for processing. As per Table 1, the average of 170,000 consumer payments are received daily in year 2021 which are equally distributed to 16 Weblogic instances for processing. It

means that each instance serves 11,000 payments daily on an average. By having this current throughput, the expected future payment traffic (load) towards each instance can be determined based on the predicted payments in 2022.

As per Table 2, it is predicted that additional payments of 17,000 are expected daily in year 2022. It means that this additional traffic is added to the current throughput of each application server instance which brings a total load of 12,100 payments (= 11,000 + 17,000/16) per each instance. In order to maintain the optimistic throughput for the payment processing, it is advised that IT operation team may plan to add at least one more Weblogic server machine (4 clusters/instances on it) to the production sandbox by end of year 2022 or beginning of 2023 for managing the 10% increased load in 2022 and henceforth.

On the other hand, this extensive payment traffic requires additional physical memory on the back-end database system (Oracle) for data repository. It is advised that IT operation team may plan to increase the physical memory size at least to 7-10 GB by end of year 2022 or beginning of 2023 for managing the 10% increased load in 2022 and henceforth.

7. References

How to Use the TimeseriesGenerator for Time Series Forecasting in Keras (2020), https://machinelearningmastery.com/how-to-use-the-timeseriesgenerator-for-time-series-forecasting-in-keras/

Mohammed Marwan (2021). Training Handouts

Time Series Prediction with LSTM Recurrent Neural Networks in Python with Keras (2020), https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networks-python-keras/