

TIME SERIES ANALYSIS AND SALES FORECASTING FOR AUTOMOTIVE USING IBM CLOUD SERVICES

1. INTRODUCTION

1.1 OVERVIEW

Sales forecasting is the process of estimating future sales with the goal of better informing your decisions. A forecast is typically based on any combination of past sales data, industry benchmarks, or economic trends. It's a method designed to help you better manage your workforce, cash flow, and any other resources that may affect revenue and sales.

Any data associated with the time that is dependent on time-related matters can be termed as time-series data. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics like trends, nonstationarity, and seasonality based on a daily, weekly, yearly basis and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

Using IBM Watson Jupyter notebook, we will perform all of the tasks involved in building predictive models for different requirements. This dataset contains two columns many columns. But we are interested in two columns for time series analysis they are month and sales to predict the sale values for different time series. The user creates an IBM Watson Studio Service, IBM Cloud Object Storage service on IBM Cloud. The user uploads sales datafile into Watson Studio. The user creates a Jupyter notebook to predict the sale value on Watson Studio. Jupyter Notebook uses Machine Learning Service to create several models, and the user deploys the best performing model. We use browser to run the application.

1.2 PURPOSE

This project aims at building a web app that automatically estimate the sale value by using the input values. It's typically easier for established businesses to create more accurate sales forecasts based on previous sales data. Newer businesses, on the other hand, will have to rely on market research, competitive benchmarks, and other forms of interest to establish a baseline for sales numbers. Using IBM Watson Jupyter notebook, we will perform all of the tasks involved in building predictive models for different requirements. You see how IBM Watson creates great models quickly, which saves time and effort, and aids in a faster decision-making process. You create a model from dataset that includes month and sales to predict the sale values for different time series.

Forecasting or Predicting the sale value helps the investors to invest in such a time where profits can be maximum. This project provides guidance to individuals who are willing to invest

or buy a car and help them in knowing the price of a day using the prophet library. It is built on the monthly sales data from 1960 - 1968. Time series analysis is made on the data for accurate predictions.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

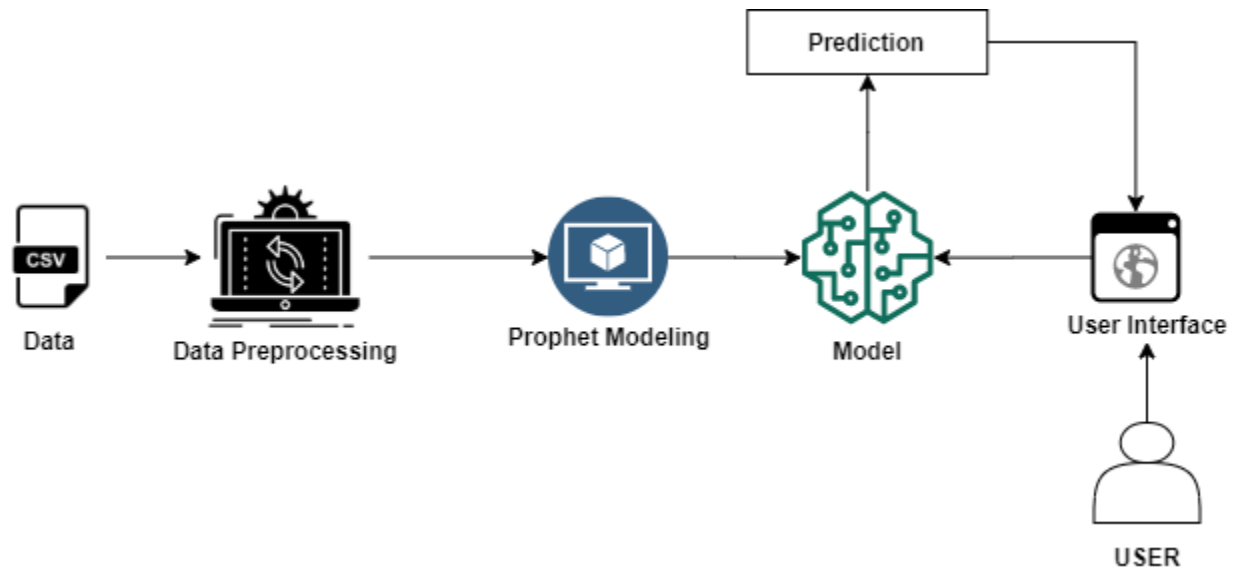
there are limitations when dealing with the unpredictable and the unknown. Time series forecasting isn't infallible and isn't appropriate or useful for all situations. Because there really is no explicit set of rules for when you should or should not use forecasting, it is up to analysts and data teams to know the limitations of analysis and what their models can support. Not every model will fit every data set or answer every question. Data teams should use time series forecasting when they understand the business question and have the appropriate data and forecasting capabilities to answer that question.

2.2 PROPOSED SOLUTION

The system being proposed here uses concept of machine learning, and the models are first trained, then tested. Finally the most accurate model will predict the final result. It first asks the date. After taking the input from the user, the system compares the input with the dataset and predicts it. It helps organizations understand the underlying causes of trends or systemic patterns over time. Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. When organizations analyze data over consistent intervals, they can also use time series forecasting to predict the likelihood of future events. In earlier we can't know the sale values. But by using this web app we can predict values from anywhere in anytime. Time series forecasting is part of predictive analytics.

3. THEORETICAL ANALYSIS

3.1 BLOCK DIAGRAM



3.2 HARDWARE / SOFTWARE DESIGNING

Hardware: Windows 10

Software:

1. Anaconda Navigator
2. Jupyter Notebook
3. Browser

IBM Watson Studio - IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud.

IBM Watson Machine Learning - IBM Watson Machine Learning helps data scientists and developers accelerate AI and machine-learning deployment.

IBM Cloud Object Storage - IBM Cloud Object Storage makes it possible to store practically limitless amounts of data, simply and cost effectively.

To build machine learning models you must require the following packages

Pandas: It is a fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language.

Numpy: It is an open-source numerical Python library . It contains a multidimensional array and matrix data structures and can be used to perform mathematical operations

fbprophet: FbProphet is a powerful time-series analysis package released by Core Data Science Team at Facebook. It is a simple and easy to go package for performing time-series analytics and forecasting at scale. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

Matplotlib: Visualisation with python. It is a comprehensive library for creating static, animated, and interactive visualizations in Python.

Seaborn: Seaborn is a library for making statistical graphics in Python. Seaborn helps you explore and understand your data. Its plotting functions operate on dataframes and arrays containing whole datasets and internally perform the necessary semantic mapping and statistical aggregation to produce informative plots

Flask: Web framework used for building web applications.

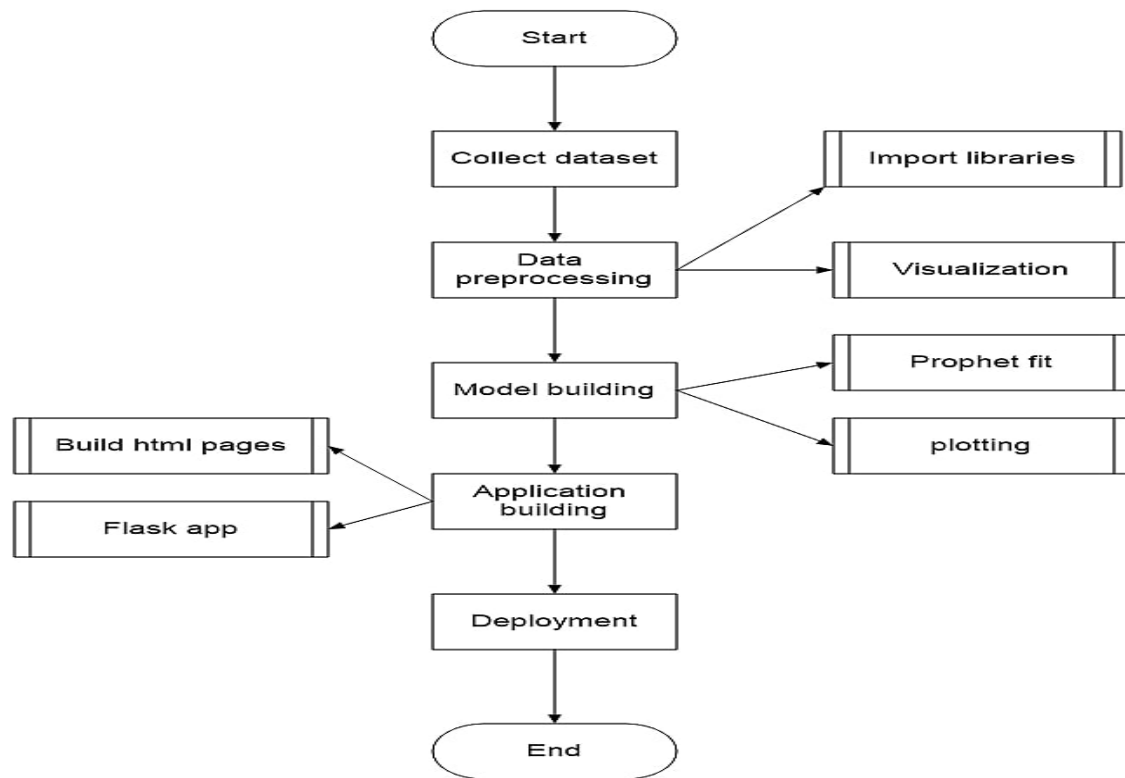
4. EXPERIMENTAL INVESTIGATIONS

Here we are going to build a machine learning model that predicts sale value prediction based on the following parameters

- month
- sales

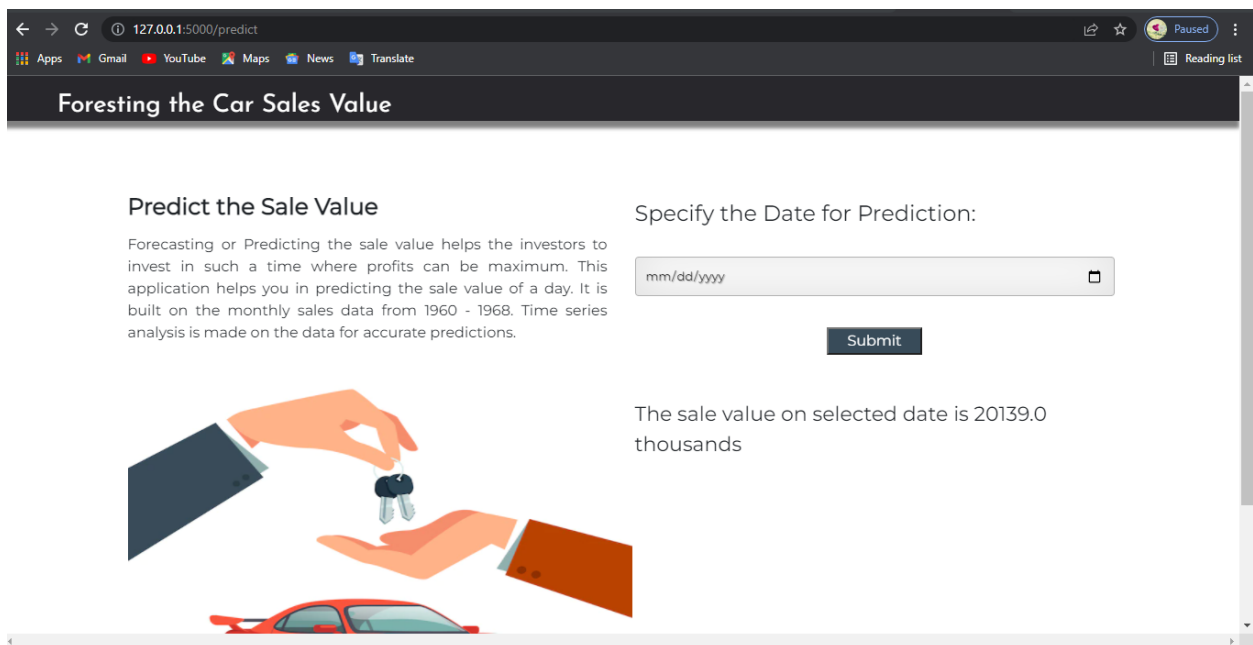
Here there are 2 parameters which affect the prediction of sale forecasting. They are month and sales. All these values are the factors which predict the expense of the customer. Having a solid sales forecast also provides a picture of your performance and performance milestones for potential investors.

5. FLOWCHART



6. RESULT

When we input the values in input field as our wish For an instance, select date as 03/8/2022



Here we got the prediction of cost of sale value as 20139.0.

7. ADVANTAGES AND DISADVANTAGES

7.1 ADVANTAGES

- Time series analysis help you to identify patterns:

The simplest and, in most cases, the most effective form of time series analysis is to simply plot the data on a line chart. With this step, there will no longer be any doubts as to whether or not sales truly peak before Christmas and dip in February.

- Create the opportunity to clean your data:

If we plotted actual sales figures for each month in the data set. If any observations were missing, the gap in the time series chart would show that right away. With any gaps in the data identified, it would be easy to impute those missing values (that is, fill in the gaps with some calculated value).

Furthermore, we would be able to identify outliers in the data. Perhaps instead of looking at actual sales, it would make more sense to plot the percentage difference between observations. This is a technique that can help smooth out very noisy data.

- Predict the future:

If we could look into a crystal ball to see the future, we would all be rich. Knowing when to expect a lull in sales, a slowdown in inventory levels, or a surge in demand would be incredibly valuable for any company.

Although not a crystal ball, time series forecasting methods can help us gain a useful glimpse of the future. While mathematically dense, the thrust of forecasting comes down to looking at past behavior and extending those patterns into the future

- Understanding data:

It can help an analyst to better understand of a dataset. This is because of the models used in time series analysis helps to interpret the true meaning of the data, as touched on previously.

- Forecasting data:

It can be the basis to forecast the data. This is because time series analysis by its very nature uncovers patterns in data, which can then be used to predict future data points.

7.2 DISADVANTAGES

- Forecast are never 100% accurate:

Let's face it: it's hard to predict the future. Even if you have a great process in place and forecasting experts on your payroll, your forecasts will never be spot on. Some products and markets simply have a high level of volatility. And in general, there is just an endless number of

factors that influence demand.

- It can be time consuming and resource intensive:

Forecasting involves a lot of data gathering, data organizing, and coordination. Companies typically employ a team of demand planners who are responsible for coming up with the forecast. But in order to do this well, demand planners need substantial input from the sales and marketing teams. In addition, it's not uncommon for processes to be manual and labor-intensive, thus taking up a lot of time. Fortunately, if you have the right technology in place, this is much less of an issue.

8. APPLICATIONS

We use this in business industries to find and get the trends of sales and investors can know the profits and losses. Your sales forecast is also your guide to how much you should be spending. Assuming you want to run a profitable business, you'll use your sales forecast to guide what you should be spending on marketing to acquire new customers and how much you should be spending on operations and administration. You're not predicting the future, but aggregating information to help define your future outlook. These assumptions are always changing, meaning that you'll need to have a pulse on the following:

9. CONCLUSION

In this project, IBM AutoAI suggested two models and evaluated the sales values. The monthly car sale data was used to develop the regression models, and the predicted sale value from these models were compared with actual values to compare the accuracies of these models. It has been found that decision tree is the best performing model. Various factors were used and their effect on the predicted amount was examined. It was observed that a the time series is affects the prediction most in every algorithm applied. Attributes which had no effect on the prediction were removed from the features.

This can help people forecasting or Predicting the sale value helps the investors to invest in such a time where profits can be maximum. This project provides guidance to individuals who are willing to invest or buy a car and help them in knowing the price of a day using the prophet library.

10. FUTURE SCOPE

While machine learning artificial intelligence may be seen as a data-hungry machine, the crucial

aspect of a successful AI system that manages a client's healthcare is its ability to develop efficient reasoning and intuitively read and understand trends. So in the future we can add more parameters to predict and analyze the sale value.

11. BIBLIOGRAPHY

- <https://www.elementai.com/news/2020/time-series-analysis-in-machine-learning>
- <https://www.geeksforgeeks.org/sales-forecast-prediction-python/>

APPENDIX

A.Source code

```
import joblib
```

```
import pandas as pd
```

```
from flask import Flask, request, render_template
```

```
from event.pywsgi import WSGIServer
```

```
import os
```

```
app = Flask(__name__)
```

```
model = joblib.load('sales.sav')
```

```
@app.route('/')
```

```
def home():
```

```
    return render_template('predict.html')
```

```
@app.route('/predict',methods=['POST'])
```

```
def y_predict():
```

```
    if request.method == "POST":
```

```
        ds = request.form["date"]
```

```
        a={"ds":[ds]}
```

```
        ds=pd.DataFrame(a)
```

```
        prediction = model.predict(ds)
```

```
        print(prediction)
```

```
        output=round(prediction.iloc[0,15])
```

```
        print(output)
```

```
        return render_template('predict.html',output="The sale value on selected date is {}
```

```
thousands".format(output))
```

```
    return render_template("predict.html")
```

```
port=os.getenv('VCAP_APP_PORT',8080)
```

```
if __name__ == "__main__":
```

```
    # app.secret_key=os.urandom(12)
```

```
    # app.run(debug=True,host='0.0.0.0',port=port)
```



```
app.run(debug=True)
```

```
import requests
```

```
# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
```

```
API_KEY = "pU7PfRmqtC1TxfZFxCZ0-QQRGvALoznaeYtXwGcGFuZ"
```

```
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
```

```
API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
```

```
mltoken = token_response.json()["access_token"]
```

```
header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
```

```
# NOTE: manually define and pass the array(s) of values to be scored in the next line
```

```
payload_scoring = {"input_data": [{"fields": [{"year", "month"}], "values": [[2005,12]]}]}
```

```
response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/4496c68a-ed82-4945-a20b-
```

```
0b33c6899d35/predictions?version=2021-12-22', json=payload_scoring,
```

```
headers={'Authorization': 'Bearer ' + mltoken})
```

```
print("Scoring response")
```

```
# print(response_scoring.json())
```

```
pred= response_scoring.json()
```

```
print(pred)
```

```
output= pred['predictions'][0]['values'][0][0]
```

```
print(output)
```