SIKSHA 'O' ANUSANDHAN

DEEMED TO BE UNIVERSITY

Admission Batch: 2020 **Session: 2020-2022**

Department of Computer Science & Engineering Faculty of Engineering & Technology (ITER)

Jagamohan Nagar, Jagamara , Bhubaneswar, Odisha – 751030

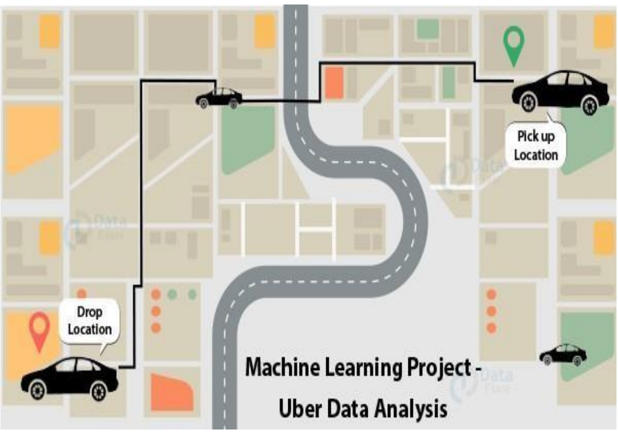
SUBMITED BY:

SWAYAM PRAVA RATH
Reg No : 2061020052

ARPITA DASH
Reg No : 2061020091

GUIDED BY:

DR. SOBHAN KANTI DHARA



UBER DATA ANALYSIS USING MACHINE LEARNING USING PYTHON [MLUP]

SUBMITTED BY:

SWAYAM PRAVA RATH:
2061020052

ARPITA DASH:
2061020091

BRANCH – M.C.A.
SEMESTER – 3RD **[B]**

UNDER THE GUIDANCE OF:

DR. SOBHAN KANTI DHARA

ABSTRACT:

Uber was founded just eleven years ago, and it was already one of the fastest-gnwing companies in the world. In Delhi , UberX claims to charge 30% less than axis a great way to get customers attention. Nowadays . we see applications of Machine Learning and Artificial Intelligence in almost all the domains so we try to use the same for Uber cahs price prediction. In this project, we did experiment with a real-world dataset and explore how machine learning algorithms could be used to find the panems in data. We mainly discuss about the price prediction of different Uber cabs that is generated by the machine learning algorithm. Our problem belongs to the regression supervised learning category. We use different machine learning algorithms, for example, Linear Regression, Decision Tree. Random Forest Regressor and Gradient Boosting Regressor but finally, choose the one that proves best for the price prediction. We must choose the algorithm which improves the accuracy and reduces overfitting. We got many experiences while doing the data preparation of Uber Dataset of Delhi of the year 2015.

KEYWORDS:

Uber , Machine Learning Using Python .

INTRODUCTION:

Looking at Data find that the data is increasing day by day and approx 2.5 quintillion bytes of data generate every day. Now, from this data analysis and get useful information which is most important and to understand that here we perform data analysis on UBER data using machine learning in Python.

Sometimes it’s easy to give up on someone else’s driving. This is less stress, more mental space and one uses that time to do other things. Yes, that’s one of the ideas that grew and later became the idea behind **Uber and Lyft**.

Both companies offer passenger boarding services that allow users to rent cars with drivers through websites or mobile apps. Whether traveling a short distance or traveling from one city to another, these services have helped people in many ways and have actually made their lives very difficult.

Uber is an international company located in 69 countries and around 900 cities around the world. Lyft, on the other hand, operates in approximately 644 cities in the US and 12 cities in Canada alone. However, in the US, it is the second-largest passenger company with a market share of 31%.

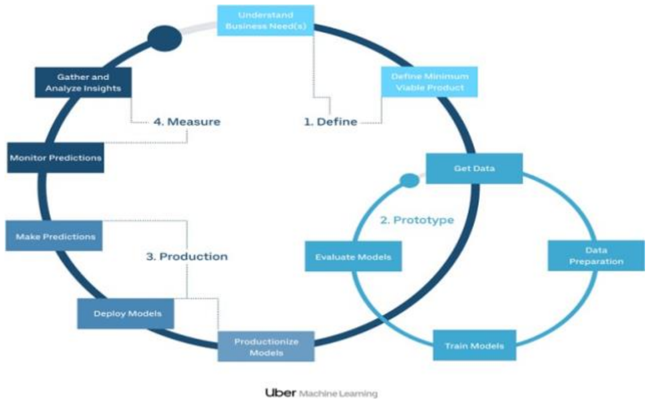
From booking a taxi to paying a bill, both services have similar features. But there are some exceptions when the two passenger services reach the neck. The same goes for prices, especially **Uber's** "**surge**" and "Prime Time" in Lyft. There are certain limitations that depend on where service providers are classified.

Many articles focus on algorithm/model learning, data purification, feature extraction, and fail to define the purpose of the model. Understanding the business model can help identify challenges that can be solved using analytics and scientific data. In this article, we go through the **Uber** Model, which provides a framework for end-to-end prediction analytics of **Uber** data prediction sources.

Overview of Uber:-

Uber's biggest competition in NYC is none other than yellow cabs, or taxis. The basic cost of these yellow cables is $ 2.5, with an additional $ 0.5 for each mile traveled. In addition, no increase in price added to yellow cabs, which seems to make yellow cabs more economically friendly than the basic UberX. However, an additional tax is often added to the taxi bill because of rush hours in the evening and in the morning. However, apart from the rising price (which can be unreasonably high at times), taxis appear to be the best option during rush hour, traffic jams, or other extreme situations that could lead to higher prices on Uber. Despite Uber's rising price, the fact that Uber still retains a visible stock market in NYC deserves further investigation of how the price hike works in real-time real estate.

Uber's Machine Learning Model :-



Workflow of ML learning project. Defining a problem, creating a solution, producing a solution, and measuring the impact of the solution are fundamental workflows. Barriers to workflow represent the many repetitions of the feedback collection required to create a solution and complete a project.

Michelangelo's "zero-to-one speed" or "value-to-one speed" is crucial to how ML spreads to Uber. In new applications, we focus on reducing barriers to entry by streamlining the workflow of people with different skills and having a consistent flow to achieve a basic model and work with good diversity.

For existing projects, we look at the speed of iteration, which shows how data scientists can quickly and get feedback on their new models or features in offline testing or online testing.

A few principles have proven to be very helpful in empowering teams to develop faster:

- Solve data problems so that data scientists are not needed.
- Dealing with data access, integration, feature management, and plumbing can be time-consuming for a data expert. Michelangelo's feature shop and feature pipes are essential in solving a pile of data experts in the head.
- Change or provide powerful tools to speed up the normal flow.
- Make the delivery process faster and more magical.
- Michelangelo hides the details of deploying and monitoring models and data pipelines in production after a single click on the UI.
- Let the user use their favorite tools with small cruft – "Go to the customer".
- Michelangelo allows for the development of collaborations in Python, textbooks, CLIs, and includes production UI to manage production programs and records.
- Enable interaction and reuse.
- Also, Michelangelo's feature shop is important in enabling teams to reuse key predictive features that have already been identified and developed by other teams.
- Guide the user through organized workflows.

End-to-end workflow:-

In the beginning, we saw that a successful ML in a big company like Uber needs more than just training good models – you need strong, awesome support throughout the workflow. We found that the same workflow applies to many different situations, including traditional ML and in-depth learning; surveillance, unsupervised, and under surveillance; online learning; batches, online, and mobile distribution; and time-series predictions. It does not mean that one tool provides everything (although this is how we did it) but it is important to have an integrated set of tools that can handle all the steps of the workflow.

1. Define

Defining a business need is an important part of a business known as business analysis. This includes understanding and identifying the purpose of the organization while defining the direction used. In addition, you should take into account any relevant concerns regarding company success, problems, or challenges.

2. Prototype

The users can train models from our web UI or from Python using our Data Science Workbench (DSW). At DSW, we support extensive deploying training of in-depth learning models in GPU clusters, tree models, and lines in CPU clusters, and in-level training on a wide variety of models using a wide range of Python tools available.

Finding the right combination of data, algorithms, and hyperparameters is a process of testing and self-replication. Going through this process quickly and effectively requires the automation of all tests and results.

3. Production

Once the working model has been trained, it is important that the model builder is able to move the model to the storage or production area. In Michelangelo, users can submit models through our web UI for convenience or through our integration API with external automation tools. Deployed model is used to make predictions.

4. Measure

Models are trained and initially tested against historical data. This means that users may not know that the model would work well in the past. But once you have used the model and used it to make predictions on new data, it is often difficult to make sure it is still working properly. Models can degrade over time because the world is constantly changing.

1. Customer Feedback ·

2. Customer Sentiments ·

3. Side Company Data ·

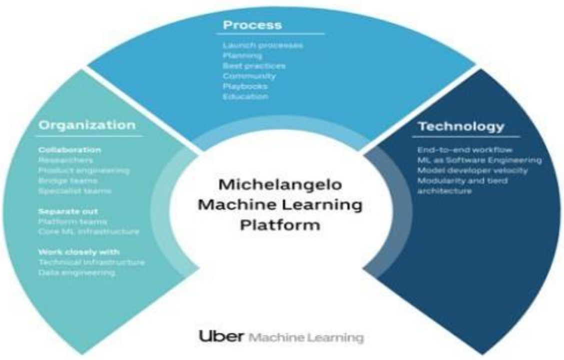
4. Situation Analysis

Requires collecting learning information for making Uber more effective and improve in the next update.

Scaling Machine Learning at Uber:-

Data scientists, our use of tools makes it easier to create and produce on the side of building and shipping ML systems, enabling them to manage their work ultimately. For developers, Uber's ML tool simplifies data science (engineering aspect, modeling, testing, etc.) after these programs, making it easier for them to train high-quality models without the need for a data scientist. Finally, for the most experienced engineering teams forming special ML programs, we provide Michelangelo's ML infrastructure components for customization and workflow.

Successfully measuring ML at a company like Uber requires much more than just the right technology – rather than the critical considerations of process planning and processing as well. In this section, we look at critical aspects of success across all three pillars: structure, process, and technology.



Organization:-

The very diverse needs of ML problems and limited resources make organizational formation very important – and challenging – in machine learning. While some Uber ML projects are run by teams of many ML engineers and data scientists, others are run by teams with little technical knowledge. Similarly, some problems can be solved with novices with widely available out-of-the-box algorithms, while other problems require expert investigation of advanced techniques (and they often do not have known solutions).

Process:-

As Uber ML's operations mature, many processes have proven to be useful in the production and efficiency of our teams. Sharing best ML practices (e.g., data editing methods, testing, and post-management) and implementing well-structured processes (e.g., implementing reviews) are important ways to guide teams and avoid duplicating others 'mistakes. Internally focused community-building efforts and transparent planning processes involve and align ML groups under common goals.

Technology:-

There is a lot of detail to find the right side of the technology for any ML system. At Uber, we have identified the following high-end areas as the most important:

End-to-end workflow: ML is more than just training models; you need support for all ML workflow: manage data, train models, check models, deploy models and make predictions, and look for guesses.

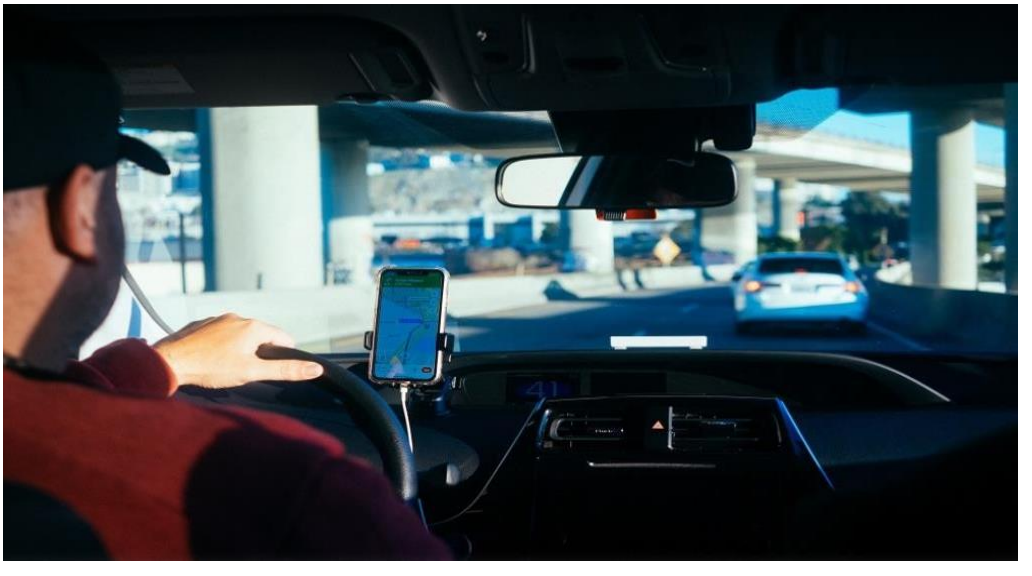
ML as software engineering: We found it important to draw analogies between ML development and software development, and then use patterns from software development tools and methods to get back to our ML functionality.

Model Developer Speed: The development of a machine learning model is a very repetitive process – new methods and advanced models come from many experiments. Because of this, the speed of the model engineers is very important.

Modularity and tiered architecture: Providing end-to-end workflow is important in managing the most common causes of ML use, but to deal with rare and very special cases, it is important to have the first things that can be integrated in a directed way.

Uber’s Dynamic Pricing Model:-

Here’s a quick and easy guide to how Uber’s dynamic price model works, so you know why Uber prices are changing and what regular peak hours are the costs of Uber’s rise.



How does Uber price work? :-

If you request a ride on Saturday night, you may find that the price is different from the cost of the same trip a few days earlier. That’s because of our dynamic pricing algorithm, which converts prices according to several variables, such as the time and distance of your route, traffic, and the current need of the driver. In some cases, this may mean a temporary increase in price during very busy times.

Why are Uber rates changing?

As demand increases, Uber uses flexible costs to encourage more drivers to get on the road and help address a number of passenger requests. When we inform you of an increase in Uber fees, we also inform drivers. If you decide to proceed and request your ride, you will receive a warning in the app to make sure you know that ratings have changed.

Price range

When more drivers enter the road and board requests have been taken, the need will be more manageable and the fare should return to normal.

Uber Top Hours

If you're a regular passenger, you're probably already familiar with Uber's peak times, when rising demand and prices are very likely. These include:

- Friday and Saturday nights
- After-work hour fast
- Major events and celebrations
- Strong prices help us to ensure that there are always enough drivers to handle all our travel requests, so you can ride faster and easier – whether you and your friends are taking this trip or staying up to you.

Business Problem:-

Before you start managing and analyzing data, the first thing you should do is think about the PURPOSE. What it means is that you have to think about the reasons why you are going to do any analysis. If you are unsure about this, just start by asking questions about your story such as **Where? What? How? Who? Which?**

Depending on how much data you have and features, the analysis can go on and on. The following questions are useful to do our analysis:

a. How many times have I traveled in the past?

b. How many trips were completed and canceled?

c. Where did most of the layoffs take place?

d. What type of product is most often selected?

e. What a measure. fare, distance, amount, and time spent on the ride?

f. Which days of the week have the highest fare?

g. Which is the longest / shortest and most expensive / cheapest ride?

h. What is the average lead time before requesting a trip?

Data visualization is certainly one of the most important stages in Data Science processes. While simple, it can be a powerful tool for prioritizing data and business context, as well as determining the right treatment before creating machine learning models.

Exploratory Data Analysis and Predictive Modelling on Uber Pickups:-

The day-to-day effect of rising prices varies depending on the location and pair of the Origin-Destination (OD pair) of the Uber trip: at accommodations/train stations, daylight hours can affect the rising price; for theaters, the hour of the important or famous play will affect the prices; finally, attractively, the price hike may be affected by certain holidays, which will increase the number of guests and perhaps even the prices; Finally, at airports, the price of escalation will be affected by the number of periodic flights and certain weather conditions, which could prevent more flights to land and land.

The weather is likely to have a significant impact on the rise in prices of Uber fares and airports as a starting point, as departure and accommodation of aircraft depending on the weather at that time. Different weather conditions will certainly affect the price increase in different ways and at different levels: we assume that weather conditions such as clouds or clearness do not have the same effect on inflation prices as weather conditions such as snow or fog. As for the day of the week, one thing that really matters is to distinguish between weekends and weekends: people often engage in different activities, go to different places, and maintain a different way of traveling during weekends and weekends. With forecasting in mind, we can now, by analyzing marine information capacity and developing graphs and formulas, investigate whether we have an impact and whether that increases their impact on Uber passenger fares .

Conclusion:-

Explanatory Data Analysis is no small feat! It takes a lot of work and patience, but it is certainly a powerful tool if used properly in the context of your business.

After analyzing the various parameters, here are a few guidelines that we can conclude. If you were a Business analyst or data scientist working for **Uber or Lyft**, you could come to the following conclusions:

- Uber is very economical; however, Lyft also offers fair competition.
- People prefer to have a shared ride in the middle of the night.
- People avoid riding when it rains.
- When traveling long distances, the price does not increase by line. However, based on time and demand, increases can affect costs.
- Uber could be the first choice for long distances.

However, obtaining and analyzing the same data is the point of several companies. There are many businesses in the market that can help bring data from many sources and in various ways to your favorite data storage.This guide briefly outlines some of the tips and tricks to simplify analysis and undoubtedly highlighted the critical importance of a well-defined business problem, which directs all coding efforts to a particular purpose and reveals key details. This business case also attempted to demonstrate the basic use of python in everyday business activities, showing how fun, important, and fun it can be.

References:-

1. Kaggle – Explanatory Data Analysis
2. Analytics Vidhya – Uber and Lyft
3. Research Gate – Explanatory Data Analysis
4. Medium (towardsdatascience) Felipe Alves Santos – Explanatory Data Analysis and Business Problem
5. Uber Blog – Dynamic pricing model
6. Uber Auth – Dataset