FOM Hochschule für Oekonomie & Management

Study Centre Essen

Master-Thesis

in degree program Big Data & Business Analytics

to obtain the degree of

Master of Science (M.Sc.)

on the subject

Impact of Artificial Intelligence in Automobile Product Development

from

Canny Sen

Supervisor	Prof. Dr. Dirk Stein MBA
Matriculation number	586465
Date of submission	16th August 2022

Table of Contents

Index of Figures

Index of Formula

List of Abbreviations

AI	Artificial Intelligence
BMC	Business Model Canvas
COA	Cost of Acquisition
ETA	Estimated Arrival Time
FNR	False Negative Rates
FRR	False Positive Rates
GAI	General Artificial Intelligence
LHC	Large Hadron Collider
LTV	Lifetime Value
ML	Machine Learning
NER	Named Entity Recognition
NLP	Natural Language Processing
PB	Petabytes
PCA	Principal Component Analysis
RL	Reinforcement Learning
SLR	Systematic Literature Review
SML	Supervised Machine Learning
SSL	Semi-supervised Learning
TB	Terabytes

1. Introduction

1.1 Importance of Topic

Unlike any other moment in history, a variety of revolutionary technologies have been introduced over the last century.[1] The automobile has influenced people deeply enough to change their culture and way of life out of these remarkable technologies. Aside from its obvious role as a mode of transportation, the vehicle has had a significant impact on one's life in a variety of unnoticed but critical ways. It has had an impact on a person's daily life in terms of the physical, economic, and social elements.

The car industry is always full of innovations and progress, especially in technology. Google Maps is used to drive over 1 billion kilometres each day in more than 220 nations and territories around the world. When people get in their car and start navigating, they're immediately given which way to go, if there's heavy or light traffic along their route, an anticipated journey time, and an estimated arrival time (ETA).

Congestion is a common occurrence in major and medium-sized cities around the world. It is an effective approach to controlling traffic congestion with limited urban facilities and resources by studying and anticipating traffic flow.[2] The predictive model is being used to assist drivers by suggesting the shortest route.

Human intelligence is replicated by machines that are programmed to think such as humans and parody their activities. As a result, artificial intelligence (AI) is the term utilized in today's technologically advanced society.[3] AI systems can then be used as a powerful tool to make faster, more rational decisions and work more efficiently. In today's world of automobiles, a version of drive assistance is being developed that employs Al to achieve customer satisfaction.

1.2 Problem Statement

Traffic congestion/jam has been a big problem in the rest of the world for decades, owing to the road's limit at a specific place or over a set length, as well as road development

[1] Cp. *Alex Covarrubias.*, The Last Automobile Revolution, 2018, pp. 81–104.
[2] Cp. *Huan Wang et al.*, Wireless Sensor Network, 2020, p. 2.
[3] Cp. *P V Ajitha and Ankita Nagra*, Artificial Intelligence in Automobile Industry, 2022, p. 504.

projects and various weather occurrences. By addressing this issue, additional delays, vehicle operating costs, fuel running costs, and emissions will be reduced.

1.3 Research Questions

- RQ1**.** How AI can holistically transform and optimize the automobile industry and its challenges?
- RQ2**.** How ML predictive models could be used to help drivers by analysing traffic data more efficiently?
- RQ3**.** What are some of the applications of AI in the automobile industry which could revolutionize the industry?

1.4 Planned Output

The main purpose of this thesis is to development of an AI/ML algorithm which will reduce factors like incremental delay, vehicle operation cost, fuel operating, and cost emission by suggesting the shortest or best possible route for drivers.

1.5 Thesis Structure

The thesis is structured as follows the standard guidelines provided by the university. The first chapter starts with an introduction to the topic which provides an explanation of the importance of the topic, defines a problem statement, research questions, and planned output of the thesis. The second chapter of the theoretical foundation presents the background and details of machine learning and Artificial Intelligence (AI) beginning from the base of big data, then explains the business model and Business Model Canvas (BMC) as well as the concepts of the automobile industry. The third chapter of the thesis explains the different analytical methods for the different types of research methodology and gives a justification approach to the thesis. This chapter also describes a case point of giving assistance to drivers using a predictive model which included a project plan, data collection, and algorithm description. The fourth and the fifth chapter present the solution to the problem statement. The first part of chapter fourth shows the data modelling, evaluation, and deployment of the thesis. The next part reflects the results and findings of the research. The fifth chapter discusses the projection of results which explains the answers to three research questions of the thesis. Lastly, the sixth chapter

discusses the conclusion drawn from the analysis and speaks about the application, challenges, and future scope of Artificial Intelligence in the automobile industry.

2. Theoretical Foundation

In this chapter, the theoretical aspects are used for reference. This chapter explained in detail Big Data analytics and their analytical methods. Furthermore, the chapter discusses Artificial Intelligence and Machine Learning from their roots in Big Data.

2.1 Overview of Big Data Analytics

Big Data term has its significance, which is applied to datasets and does not conform to the normal structure of the traditional database.[4] They are datasets that are too large for frequently used software tools and storage systems to gather, store, manage and process within a reasonable amount of time.[5]

The scale of big data is constantly growing, with a single dataset now ranging from a few dozen terabytes (TB) to many petabytes (PB).[6] Humans created 5 exabytes (1018 bytes) of data till 2003, this much data is now created in just a couple of days. According to the reports of Statista, the overall data generated by the end of 2021 was around 74 zettabytes (1 zettabyte = 1000 exabytes, 1 exabyte = 1000 petabytes, and 1 petabyte = 1000 terabytes roughly) and is enlarged to more than 180 zettabytes by the year 2025.[7]

Figure 1: Definition of Big Data Analytics

Sr. No.	Definition of Big Data Analytics	Author
1.	Big data is a term that is used to describe data that is high volume, high velocity, and/or high variety; requires new technologies and techniques to capture, store, and analyze it; and is used to enhance decision making, provide insight and discovery, and support and optimize processes.	Mills et al
2.	Big data is where the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant scaling (more nodes) for efficient processing.	National Institute of Standards and Technology, USA
3.	Big data is a term describing the storage and analysis of large and/or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce, and machine learning.	Ward and Barker
4.	A data set that is too big to fit on a screen. Data sets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze	Shneiderman (2008), Manyika et al. (2011)
5.	Data that cannot be handled and processed in a straightforward manner. The data sets and analytical techniques in applications that are so large and complex that they require advanced and unique data storage, management, analysis, and visualization technologies.	Fisher et al. (2012), Chen et al. (2012)

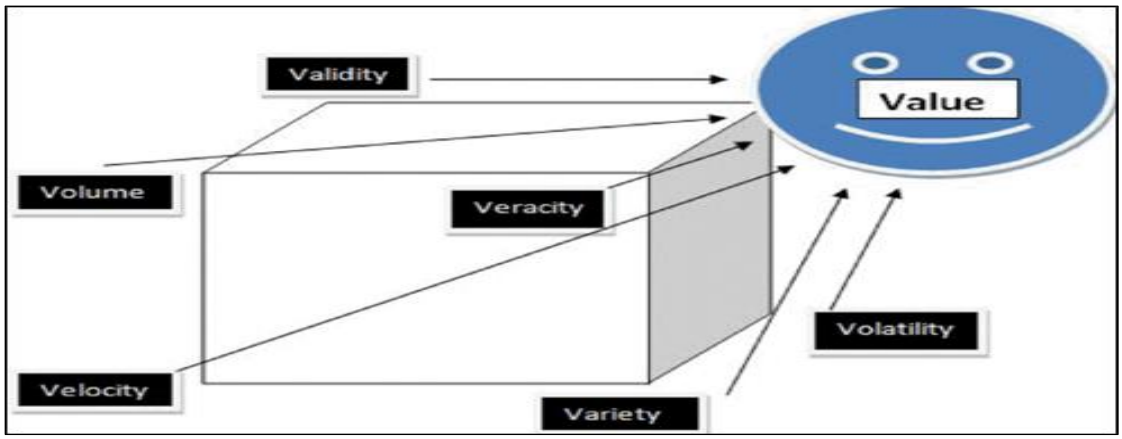
Source: Author's own representation

[4] Cp. *Nada Elgendy and Ahmed Elragal.*, Big Data Analytics, 2014, pp. 214–27.
[5] Cp. *Seref Sagiroglu and Duygu Sinanc.*, Big Data, 2013, pp. 42–47.
[6] Cp. *Mary Anne M. Gobble.*, The Next Big Thing, 2015, pp. 64–66.
[7] See https://www.statista.com/statistics/871513/worldwide-data-created/, accessed 14.06.2022; *No Author*, Total Data Volume Worldwide, no page number.

The term "big data" refers to a large amount of information. The term "size" is a relative one.[8] 20 Megabytes was considered huge in the 1960s. Data is no longer considered large until it reaches many hundred petabytes. Other than volume (size), there are several other properties that are commonly discussed.

The 7 Vs of big data are as follows:

Figure 2: 7 Vs of Big Data



Source: Based on https://silo.tips/download/seven-v-s-of-big-data, accessed 15.06.2022

Volume: The quantity of information generated from all sources, including text, audio, video, social networking, research studies, medical data, space photos, criminal reports, weather predictions, natural disasters, and so on is referred to as Big data volume.[9] For instance, folks on Facebook alone, send around 10 billion messages per day, every day the average user likes about 4.5 billion posts and uploads 350 million new photos. However, due to the unstructured and unknown nature of the data, it cannot be managed, processed, or queried using typical methods such as SQL. It needs special and powerful tools like Kafka, Hadoop, etc.[10]

Velocity: When it comes to vast and complex data, this is another key term to discuss. This is the data's speed or velocity which makes it difficult to deal with efficiently.[11] Within and outside enterprises, the speed with which data is moved, processed, and captured has increased dramatically. Business intelligence models often take days to process, however, today's analytical demands necessitate data gathering and processing

[8] Cp. *Muhammad Naeem et al.*, Trends and Future Perspective Challenges, 2022, pp. 309–25.
[9] Cp. *Amir Gandomi and Murtaza Haider*., Big Data Concepts, 2015, pp. 137–44.
[10] Cp. *Maryam M. Najafabadi et al.*, Deep Learning Applications and Challenges, 2015, pp. 1–21.
[11] Cp. *M. Ali Ud Din Khan., Muhammad Fahim Uddin, and Navarun Gupta*., Seven V's of Big Data, 2014, no page number.

'practically' in real-time due to the high-speed flow of data.[12] Phone calls, data collected from experiments, data sent by sensors, data transferred over the Internet, and stock price data are all examples of real-time data. Because large amounts of data are temporary, they must be examined as they are generated. They quickly become obsolete.

Variety: Image, audio, and video data nowadays are available in addition to numbers and text. The data generated by the Large Hadron Collider (LHC), as well as earth and polar observations, is mostly numerical. Unstructured textual data is generated mostly through word processors, emails, tweets, blogs, and other social media.[13] Image data includes medical pictures and billions of photographs taken using mobile phones. Video data is generated by surveillance cameras and movies. Audio data is stored on music websites. In the 1980s, most data were structured and organized as tables with keys. They are now unstructured and multimedia data are frequently combined. According to estimates, 90% of today's data is generated in an unstructured fashion. And because not every data can be analysed with every method, such methods must adapt to the nature of the data.

Veracity: Before this V has been added to big data, incoming data is clean and precise which is considered by the scientific and research sector. However, V is of the utmost importance to the processing of large data, as well as related analysis and findings.[14] Veracity means the truthfulness of data.[15] Masses can trust the data, and there are no duplicates. As a result, it is critical to think about cleansing massive data with certain excellent tools and algorithms.

Validity: Validity of data and veracity of data may have similar principles; however, they may not share the very same concepts and theories. Data should be legitimate when it transitions from the exploratory to actionable stage.[16] A data set may not have any veracity issues, yet it may not be genuine if it is difficult to comprehend. This Big Data feature is critical for detecting the presence of hidden links among items in large Big Data generating sources. Likewise, a set of data may be verified for one application or usage but invalid for another.

---

[12] Cp. *Thibaud Chardonnens et al.*, High Velocity Streams, 2013, pp. 784–87.
[13] Cp. *Anto Praveena and B. Bharathi*., Big Data Analytics, 2017, no page number.
[14] Cp. *Nawsher Khan et al.*, The 10 Vs, 2018, no page number.
[15] Cp. *Khan, Uddin, and Gupta*, Big Data to Extract Value, no page number.
[16] Cp. *Khan et al.*, The 10 Vs, no page number.

Volatility: The term "big data volatility" relates to the data's life span, or how much it is valid and how far it should be stored. One must identify at what point and when real-time data is no longer relevant and appropriate to the current study in this field. The data should always be present in some sources, while this may not be the case in others.[17] As a result, this is required in order to comprehend data requirements, availability, and lifetime. Because of the amount, pace, and variety of big data, understanding volatility is critical. For certain sources, the same data will always be available; however, this wouldn't be the case for others. Understanding what data is available and for how long can assist people in defining big data retention needs and regulations.

Value: Data value refers to the utility of data in a decision-making and is among the most important aspects of Big Data since it has a major effect on company earnings.[18] The key to Big Data is not how much data there is, but how that data is used and/or handled. Even while implementing IT infrastructures to manage enormous volumes of data is costly, it may provide organizations with significant competitive advantages.[19]

2.2 Analytical Methods for Big Data Analytics

The extraction of usable knowledge and insights from huge data is the focus of data analytics.[20] This is accomplished by formulating hypotheses, which are frequently based on speculations obtained from experience, and detecting connections between variables.[21] Although there are numerous ways to analyse big data, the common ones that are frequently used are Descriptive, Diagnostic, Predictive, and Prescriptive.

Figure 3: Different types of Analytical Methods

[17] Cp. *ibid*
[18] Cp. *Khan, Uddin, and Gupta.*, Seven V's of Big Data, no page number.
[19] Cp. *Ishwarappa and J. Anuradha.*, A Brief Introduction, 2015, pp. 319–24.
[20] Cp. *Uma Narayanan, Varghese Paul, and Shelbi Joseph.*, Different Analytical Techniques, 2018, pp. 372–82.
[21] Cp. *Uthayasankar Sivarajah et al.*, Critical Analysis of Big Data, 2017, pp. 263–86.

Source: Based on https://hevodata.com/learn/ultimate-guide-to-data-analytics/, accessed 15.06.2022

Descriptive Analytics: This essentially uses data aggregation and data mining of what happened in the past to give insight and does so in an understandable manner.[22] Data is visualized using bar charts, graphs, pie charts, maps, scatter diagrams, and other visualization to provide insight as to what the data means. This type of data presentation is sometimes referred to as a dashboard because it resembles the dashboard of an automobile, which displays information such as speed, engine status, fuel remaining in the tank, distance travelled, and so on.[23]

The analysis can be done using normal programming languages like Python, R, etc. MS Excel and Power BI are the tools that can be used to generate reports on this analysis. There are various techniques described in descriptive statistics such as Measures of central tendency (Mean, Median and Mode), Measure of Position (Percentiles), and Measure of Variation (Standard Deviation, Variance, and Skewness). [24] [25]

Figure 4: Representation of Descriptive Statistics

Sr. No.	Brand/model	Mean	Median	Max.	Min.	SD	Skewness	Kurtosis
0	Ford SUV	1028.17	1028	3400	66	696.86	1.16	4.58
1	Honda SUV	1132.11	1074	2716	185	483.25	0.73	3.83
2	Mitsubishi 2.0	1088.21	917	3031	121	645.3	1.09	4.24
3	Nissan SUV	1177.09	1093	4132	51	865.1	1.27	4.95
4	Suzuki 1.3	647.26	676	1519	85	409.23	0.12	1.78
5	Toyota 3.0	183.62	151	579	40	134.87	1.62	5.1

Source: Author's own representation

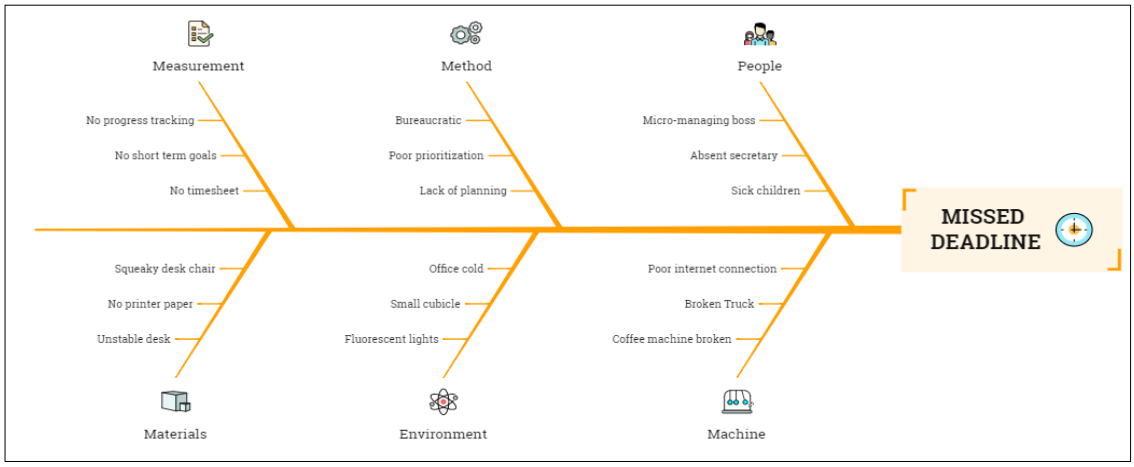
Diagnostics Analytics: Diagnostic analytics, like descriptive analytics, relies on previous data to respond to a query. Diagnostic analytics, on the other hand, instead of focusing on "what," asks why an occurrence or anomaly occurred in the data.[26] Drill-down, Data Discovery, Data Mining, and Correlations are examples of strategies that take a more in-depth look at data to try to understand the reasons for occurrences and behaviours.[27] Data from Facebook, emails, sales patterns, and so forth. Organizations will be able to predict

[22] Cp. *Harkiran Kaur and Aanchal Phutela*., Descriptive Data Analytics, 2018, pp. 678–83.
[23] Cp. *Hung Cao, Monica Wachowicz, and Sangwhan Cha.,* Real-Time Descriptive Analytics, 2017, pp. 4546–54.
[24] Cp. *Graham Williams*., Descriptive and Predictive Analytics, 2011, pp. 171–77.
[25] Cp. *Wei Chun Hsu, Lin Lin, and Chen Yu Li.*, Forecasting Automobile Sales, 2014, pp. 568–80.
[26] Cp. *Fakhitah Ridzuan and Wan Mohd Nazmee Wan Zainon*., Diagnostic Analysis, 2021, pp. 685–92.
[27] Cp. *Tiko Iyamu.*, Information Management Affiliation, 2020, no page number.

client activities such as renewing magazine subscriptions, changing mobile phone service providers, and cancelling hotel reservations based on their behaviour. A corporation may then make an enticing offer in an attempt to sway the user's expected behaviour.

Diagnostic analytics, also known as root cause analysis, is used to discover the source of business issues and find effective remedies to avoid them from recurring in the long term.[28]

Figure 5: Representation of Diagnostics Analytics



Source: Author's own representation

Predictive Analytics: Predictive analytics is a type of advanced analytics that uses machine learning to predict what will happen based on prior data. Predictive analytics models are built using historical data, which makes up the majority of descriptive and diagnostic analytics.[29] It's crucial to keep in mind that no statistical technique can accurately anticipate the future. From forecasting consumer behaviour and purchase patterns to finding trends in sales activities, predictive analytics can be applied throughout the enterprise.[30] Time series analysis using statistical approaches, neural networks, and machine learning algorithms are employed for extrapolation.[31]

Different metrics are used to perform predictive analytics such as False Positive Rates (FPR), False Negative Rates (FNR), R-Square, Odds Ratio, Recall, Precision, etc.

---

[28] See https://www.researchgate.net/figure/Root-cause-analysis_fig7_264836341, accessed 15.06.2022; *No Author*, Root Cause Analysis, 2022, no page number.
[29] Cp. *Matthew A. Waller and Stanley E. Fawcett*, Predictive Analytics, 2013, pp. 77–84.
[30] Cp. *Galit Shmueli and Otto R. Koppius*, Information Systems Research, 2011, pp. 553–72.
[31] Cp. *Vaibhav Kumar and M. L*, Trends and Techniques, 2018, pp. 31–37.

Figure 6: Representation of Metrics for Predictive Analytics

	precision	recall	f1-score	support
0	0.88	0.45	0.60	51
1	0.82	0.98	0.89	134
accuracy			0.83	185
macro avg	0.85	0.71	0.75	185
weighted avg	0.84	0.83	0.81	185

Source: Author's own representation

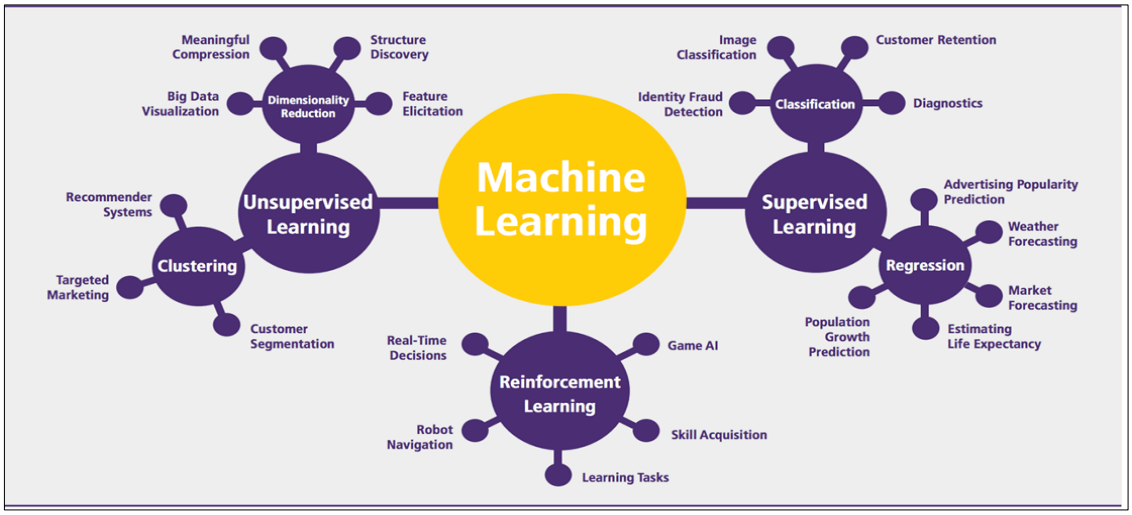
Prescriptive Analytics: The last pillar of modern analytics is prescriptive analytics. Prescriptive analytics refers to true directed analytics, in which the data is prescribing or directing users to take a specific action. To drive decision-making, it effectively combines descriptive and predictive analytics. Existing situations or conditions, as well as the consequences of a decision or event, are used to generate a user-guided decision or action.[32] To put it another way, the analysis teaches everyone how to reach a goal.

2.3 Machine Learning Methodologies for Big Data Analytics

Machine learning is a branch of science that studies the theory, performance, and properties of learning algorithms and systems.[33] It is considered Artificial Intelligence, and it automatically learns and improves performance by trial and error method.[34] There are applications in which the use of machine learning is very common. Netflix's movie recommendations, Facebook's friend tagging, Amazon's Alexa, and even Gmail's spam filtering are all examples of applications that use machine learning algorithms.[35] Supervised Learning, Unsupervised Learning, and Reinforcement Learning are the three types of machine learning approaches in the field of Big Data Analytics.[36]

[32] Cp. *Katerina Lepenioti et al.*, Prescriptive Analytics, 2020, pp. 57–70.
[33] Cp. *Junfei Qiu et al*., Machine Learning for Big Data Processing, 2016, pp. 1–16.
[34] Cp. *Eugenio Parra et al*., Quality of Requirements, 2015, pp. 180–95.
[35] Cp. *Shweta Mittal and Om Prakash Sangwan*, Machine Learning Techniques, 2019, pp. 203–7.
[36] See https://medium.com/analytics-vidhya/which-machine-learning-algorithm-should-you-use-by-problem-type-a53967326566, accessed 21.06.2022; *Sukanya Bag*, 2022, Machine Learning Algorithm, no page number.

Figure 7: Representation of Machine Learning Algorithms



Source: Based on https://medium.com/analytics-vidhya/which-machine-learning-algorithm-should-you-use-by-problem-type-a53967326566, accessed 22.06.2022

2.3.1 Supervised Machine Learning Algorithms

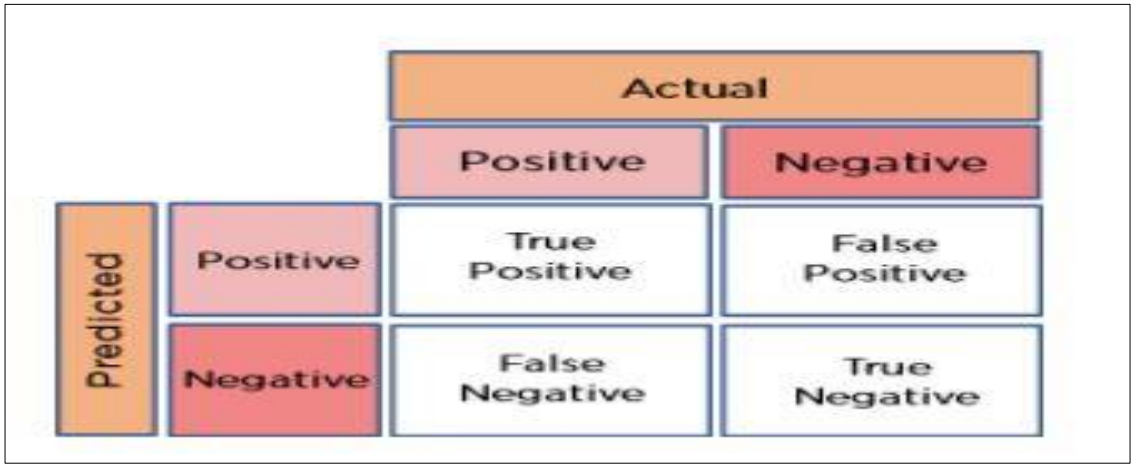
Supervised Machine Learning (SML) is an algorithm that learns from labelled training data to generate hypnosis, which then forecasts occurrences in the future.[37] It requires time and technical know-how from a team of qualified data scientists to create, scale, and deploy accurate supervised machine learning models successfully. Additionally, data scientists must re-create models to ensure that the insights provided are accurate until the data changes.[38] The method of supervised learning involves giving the machine learning model both the right input data and output data. Finding a mapping function to link the input variable (x) with the output variable (y) is the goal of a supervised learning algorithm.[39]

Supervised Machine Learning Algorithms are classified into two types called Classification Algorithm and Regression Algorithms.[40] On the availability of training data, the classification algorithm is a supervised learning method that is used to categorize new observations. In classification, a program makes use of the dataset or observations

[37] Cp. *Akinsola Jet and Hinmikaiye J O.*, Supervised Machine Learning Algorithms, 2017, no page number.
[38] Cp. *Tammy Jiang, Jaimie L. Gradus, and Anthony J. Rosellini.*, Supervised Machine Learning, 2020, pp. 675–87.
[39] Cp. *Crisci, B. Ghattas, and G. Perera.*, Supervised Machine Learning Algorithms, 2012, pp. 113–22.
[40] Cp. *Amanpreet Singh, Narina Thakur, and Aakanksha Sharma.*, Supervised Machine Learning, 2016, pp. 1310–15.

that are provided to learn how to categorize fresh observations into various classes or groups. For instance, cow or buffalo, yes or no, 0 or 1, and so on.[41] There are several different classification algorithms like Decision Trees, Random Forest, Naïve Bayes, etc.[42] The primary objective of a classification algorithm is to determine the category of a given dataset, and these algorithms are primarily employed to forecast the results for categorical data.[43] The algorithms are created classification on a dataset is called a classifier.[44] There are two categories of a classifier. The first one is a binary classifier which is used when there are only two feasible solutions to a classification problem. To exemplify Yes or No, Male or Female, Spam or Not Spam, and so on. Another one is a multi-class classifier in which a classification problem is referred to as multi-class if there are more than two possible results. For example, classifications of various crop types and musical genres. The confusion matrix describes the performance of the model and gives us a matrix or table as an output.

Figure 8: Representation of Confusion Matrix



Source: Based on https://www.simplilearn.com/tutorials/machine-learning-tutorial/confusion-matrix-machine-learning, accessed 22.06.2022

Regression analysis is a statistical approach for looking at how different variables relate to one another.[45] Regression analysis uses one or more independent variables to describe

[41] Cp. *Iqbal H. Sarker*., Applications and Research Directions, 2021, pp. 1–21.
[42] See http://sersc.org/journals/index.php/IJAST/article/view/27585, accessed 24.06.2022; *No Author*, Frequently Used Classification Algorithms, 2022, no page number.
[43] Cp. *J. Gama and P. Brazdil*., Classification Algorithms, pp. 189-200.
[44] Cp. *L. B. Booker, D. E. Goldberg, and J. H. Holland*., Classifier Systems and Genetic Algorithms, 1989, pp. 235–82.
[45] Cp. *Alan O Sykes*, Regression Analysis, 2022, no page number.

the relationship between a dependent (target) and independent (predictor) variables.[46] More specifically, regression analysis enables us to comprehend how, while other independent variables are held constant, the value of the dependent variable changes in relation to an independent variable. This forecasts real, continuous values like temperature, age, salary, and cost, among others.

Regression displays a line or curve that crosses through each data point on the target-predictor graph in such a way that the vertical separation between the data points and the regression line is minimized.[47] Whether a model has captured a strong relationship or not can be inferred from the distance between the data points and line.

A continuous variable can be predicted with the aid of regression analysis. In actual life, there are many situations where the masses need to make predictions about the future, including those involving the weather, sales, marketing trends, and other factors.[48] In these situations, one needs technology that can make forecasts more precisely.

2.3.2 Un-supervised Machine Learning Algorithms

A machine learning technique called unsupervised learning uses training datasets without supervising the models. Rather, models themselves decipher the provided data to reveal hidden patterns and insights.[49] It is comparable to the learning process that occurs in the human brain while learning something new.[50]

Unlike supervised learning, where input data is available but no corresponding output data, unsupervised learning cannot be applied directly to a regression or classification task.[51] Finding the internal structure of a dataset, classifying the data into groups of commonalities, and representing the dataset in a compressed format are the objectives of unsupervised learning.[52]

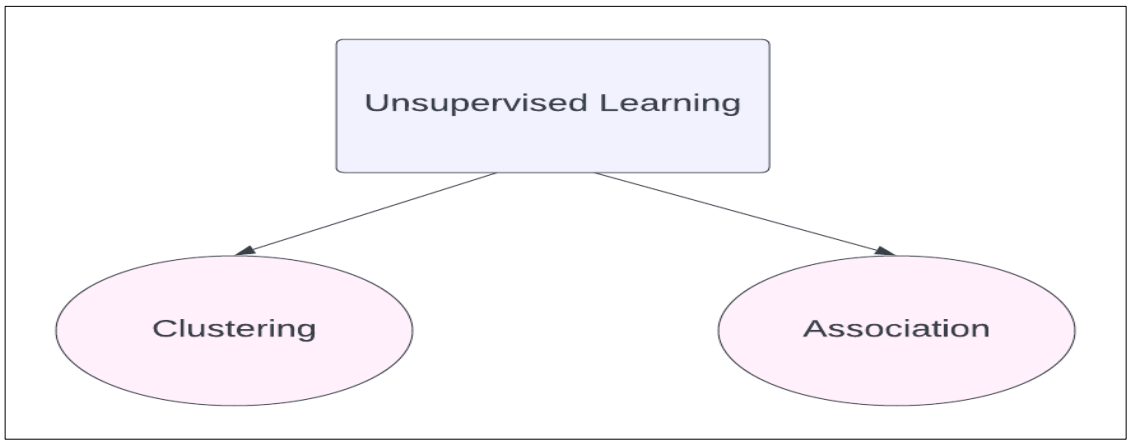
---

[46] Cp. *Mittal and Sangwan*., Machine Learning Techniques, no page number.
[47] Cp. *Sykes.*, An Introduction to Regression Analysis, no page number.
[48] Cp. *Gülden Kaya Uyanık and Neşe Güler.*, Multiple Linear Regression Analysis, 2013, pp. 234–40.
[49] Cp. *Memoona Khanum Tahira Mahboob Assistant Professor Assistant Professor and Warda Imtiaz Humaraia Abdul Ghafoor Rabeea Sehar*., Unsupervised Machine Learning Algorithms, 2015, pp. 975–8887.
[50] Cp. *Juan Carlos Niebles et al*., Unsupervised Learning, 2008, pp. 299–318.
[51] Cp. *Mohamed Alloghani et al*., Supervised and Unsupervised Machine Learning Algorithms, 2020, pp. 3–21.
[52] Cp. *Madhumita Sahoo*., Modeling Approaches, 2022, pp. 87–103.

Some of the most common unsupervised machine learning algorithms are K-Means clustering, PCA (Principal Component Analysis) Models, Stochastic Neighbors, etc. Popular programming languages such as Python, R, etc. can be used to carry out unsupervised algorithms.

The unsupervised learning algorithm can be divided into two different categories of issues: clustering and association.

Figure 9: Representation of Unsupervised Learning



Source: Author's own representation

Using the clustering technique, items are grouped into clusters so that those who share the most similarities stay in one group and share little to none with those in another.[53] On the other hand, an unsupervised learning technique called an association rule is used to uncover the connections between variables in a sizable database. It establishes the group of items that cross in the collection.[54]

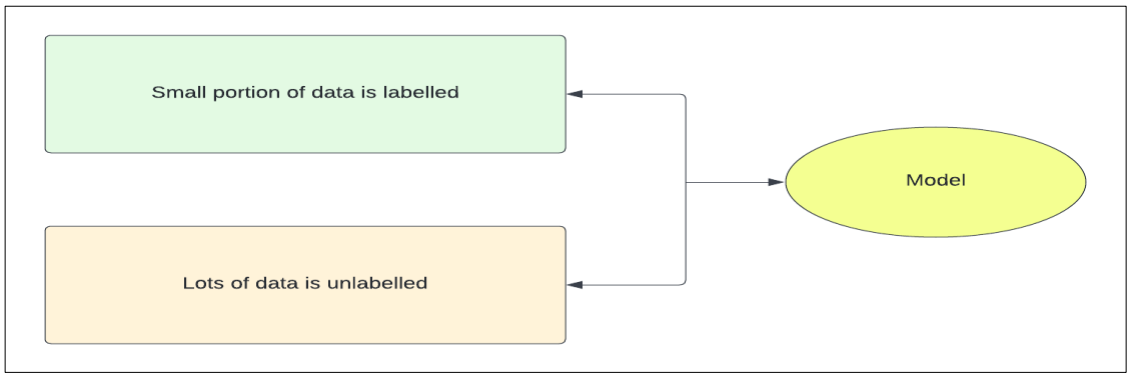
2.3.3 Semi-Supervised Machine Learning Algorithms

Semi-supervised learning (SSL) is in middle between supervised and unsupervised learning.[55] The algorithm is given unlabelled data as well as some supervision data, though not always for all examples. These types of learning issues are difficult because neither supervised learning algorithms nor unsupervised learning algorithms can

[53] Cp. *Gunjan K. Gupta, Alexander Strehl, and Joydeep Ghosh.*, Distance Based Clustering, 1999, pp. 759–64.
[54] Cp. *Feng Hsu Wang and Hsiu Mei Shao.*, Time-Framed Navigation Clustering, 2004, pp. 365–77.
[55] Cp. *Viswanath Pulabaigari et al.*, Semi-Supervised Learning, 2018, pp. 81–85.

effectively handle combinations of labelled and untellable data.[56] Specialized semi-supervised learning methods are therefore necessary.

Figure 10: Representation of Semi-supervised Learning



Source: Author's own representation

Sci-kit Learn in Python and packages in R provide the means to carry out the semi-supervised machine learning algorithms.[57] Several assumptions are made for the semi-supervised machine learning algorithms such as:

- Continuity Assumption: where-in if the independent variables have close values concerning each other they are labelled as one.
- Clustering Assumption: where-in if the data points of the independent variables fall in the same cluster, they are more likely to have the same label.
- Manifold Assumption: the data points with a similar lower-dimensional space are assumed to have the same label.

2.3.4 Reinforcement Learning

A machine learning method known as Reinforcement Learning (RL) uses feedback to teach an entity how to operate in a given environment by having it perform actions and observe the outcomes.[58] The agent receives compliments for each positive activity and is penalized or given negative responses for each negative action.[59]

In contrast to supervised learning, reinforcement learning uses feedback to autonomously train the agent without the use of labelled data. The agent can only learn from their own

[56] Cp. *Yu Feng Li and De Ming Liang*., Safe Semi-Supervised Learning, 2019, pp. 669–76.
[57] Cp. *Jesper E. van Engelen and Holger H. Hoos*., Semi-Supervised Learning, 2020, pp. 373–440.
[58] Cp. *Csaba Szepesvári*., Reinforcement Learning, 2010, pp. 1–89.
[59] Cp. *Peter Dayan and Yael Niv*., Reinforcement Learning, 2008, pp. 185–96.

experience because there isn't any labelled data.[60] In situations that include gaming, robotics, and other long-term endeavours where decisions must be made sequentially, RL offers a solution.[61]

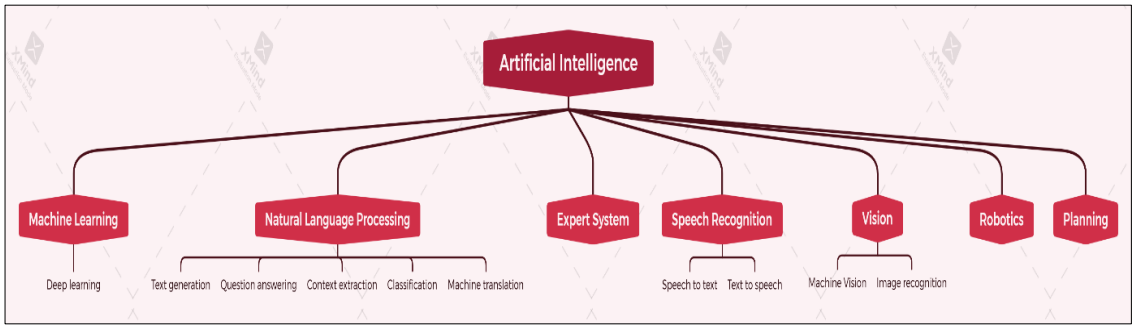
Reinforcement Learning can be performed using python packages such as Pyqlearning and also is used to carry out deep reinforcement learning. One of the examples of deep reinforcement learning algorithms is Keras-RL.

2.4 Artificial Intelligence

Machines are becoming smarter every day and are being enhanced with embedded intelligence as part of the present industrial revolution led by artificial intelligence in order to undertake intensive data analysis, find the source of a problem, and produce insightful information.[62]

The ability of machines to perform and display human-like behaviour and intellect, such as solving challenging problems without precise instructions incorporated in the software coupled with particular action orders, is known as artificial intelligence (AI).[63] Artificial intelligence is already widespread in our world. It is currently engaged in a wide range of subfields, from the general to the specialized, including self-driving cars, chess play, theorem proving, music performance, painting, etc.[64] The public and the scientific community are equally fascinated by AI's recent successes. An illustration of this is the capability of autonomous vehicles to make rational decisions on drives in a variety of real-traffic road circumstances.

Figure 11: Sub-sections of Artificial Intelligence



[60] Cp. *yuxi Li.*, Deep Reinforcement Learning, 2018, no page number.
[61] Cp. *Rashmi Sharma, Manish Prateek, and Ashok K. Sinha.*, Reinforcement Learning, 2013, pp. 28–34.
[62] Cp. *Rahul Reddy Nadikattu*, Role Of Artificial Intelligence, 2016, pp. 2320–2882.
[63] Cp. *Nils J. Nilsson.*, Artificial Intelligence, 1996, pp. 369–80.
[64] Cp. *Russell S and Bohannon J.*, Artificial Intelligence, 2015, pp. 252–252.

Source: Based on https://www.researchgate.net/publication/342106972_The_role_of_Artificial_Intelligence_in_future_technology, accessed 01.07.2022

2.4.1 A brief history of Artificial Intelligence

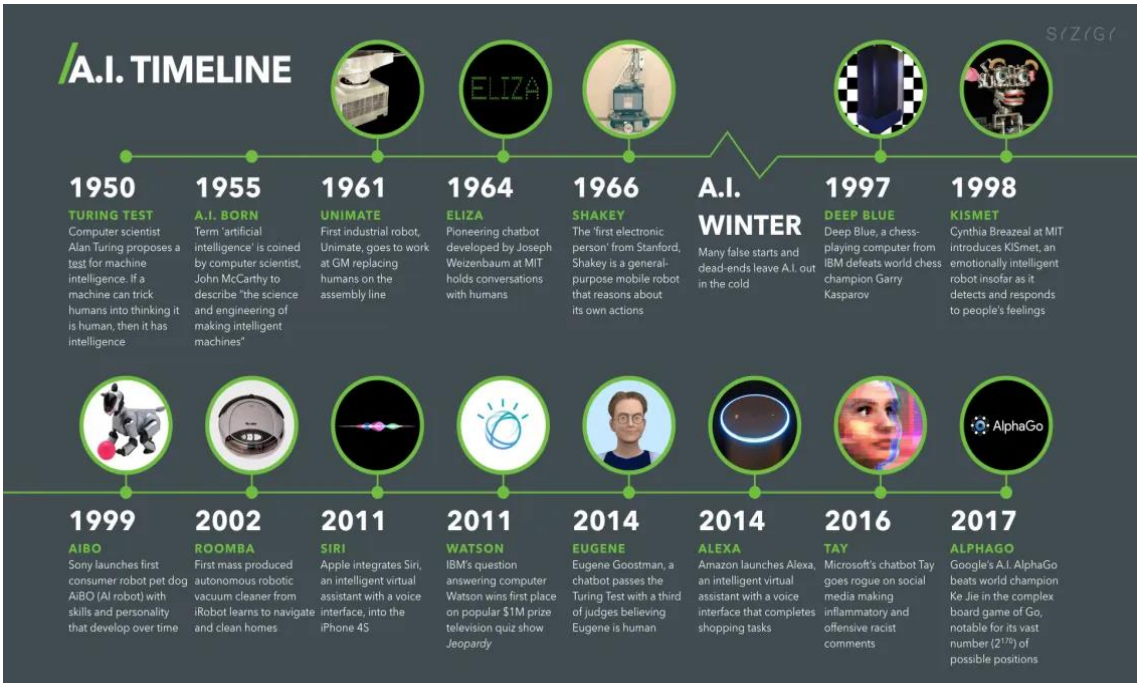
Early advancements in computers, chemistry, and a number of other disciplines have influenced AI. Early inventions included work such as problem-solving, which required foundational work in understanding, knowledge representation, and inference.[65] Furthermore, early milestone attempts encompassed language comprehension, translation, theorem proving, and evidence-based frameworks.

John McCarthy initially used the phrase artificial intelligence in 1956 when he hosted the first scholarly conference on the topic.[66] However, the quest to discover whether or not machines can actually think began far earlier. Vannevar Bush developed a system that magnifies people's own knowledge and comprehension in his seminal book As We May Think.[67] After 5 years, Alan Turing published a paper on the idea that machines could mimic humans and be intelligent enough to perform tasks like playing chess.[68]

In 1974, Marvin Minsky proposed an important information representation theory, and in 1977, Edward Feigenbaum released "The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering". However, from 1970 until the mid-1990s, many scientists and researchers had a funding crisis for AI study, owing to an immense volume of data, and this period became known as the ‘AI Winters’.[69]

[65] Cp. *Attila Benko and Cecília Sik Lányi*., History of Artificial Intelligence, pp. 1759–62.
[66] Cp. *Scott L. Andresen*., Father of Al, 2002, pp. 84–85.
[67] See http://www2.theatlantic.com/, accessed July 3, 2022; *Vannevar Bush*, 2022, As We May Think, no page number.
[68] Cp. *Stephen Muggleton*, Development of Artificial Intelligence, 2014, pp. 3–10.
[69] Cp. *Yanqing Duan, John S. Edwards, and Yogesh K. Dwivedi.*, Decision Making, 2019, pp. 63–71.

Figure 12: Representation of Artificial Intelligence Timeline



Source: Based on https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/, accessed 01.07.2022

AI was thriving in the late 1990s, and it supported various countries in building and implementing the fifth Generation Computer System.[70] Japan and China developed around 863 National High-Tech Programs and aided other developing countries in the establishment of the first AI research centres.

2.4.2 Definition of Artificial Intelligence

In general, since Al is a concept that spans and combines a broad range of technologies and the large quantities of application sectors are making a clear and unique definition much more difficult.[71] Going to define AI as some particular technology is challenging for a variety of reasons. While some of these technologies and application fields are game-changing and disruptive, others just bring slight enhancements to previously used methods, procedures, and practices.[72] The lack of a universal theory of what and how one defines 'intelligence', and what it is actually composed of, presents additional challenges related to the AI definition.

[70] Cp. *J Paul Myers*, Computing Project, 2021, pp. 77–86.
[71] Cp. *Pei Wang*., Defining Artificial Intelligence, 2019, pp. 1–37.
[72] Cp. *Ibid.*

Regardless of the debate over the exact definition, the heart of Al consists of a series of algorithms that either follow a set of instructions or a set of restrictions. Combining the algorithms yields an AI model that, using predetermined rules and probabilities, derives the best possible output from any given input.[73] An explanation is necessary to better prepare for the understanding of the impact AI can have, even when the technology component of Al is difficult to grasp.

Burgess (2017) divides artificial intelligence into supervised and unsupervised categories with a focus on how it works, but Barrat (2013) discusses three categories of AI depending on the level of intelligence: artificial narrow intelligence, artificial general intelligence, and artificial super intelligence.[74] The last two of Barrat's (2013) concepts are necessary to grasp but are merely speculative predictions and theorizations of what AI might become.

2.4.3 Types of Artificial Intelligence

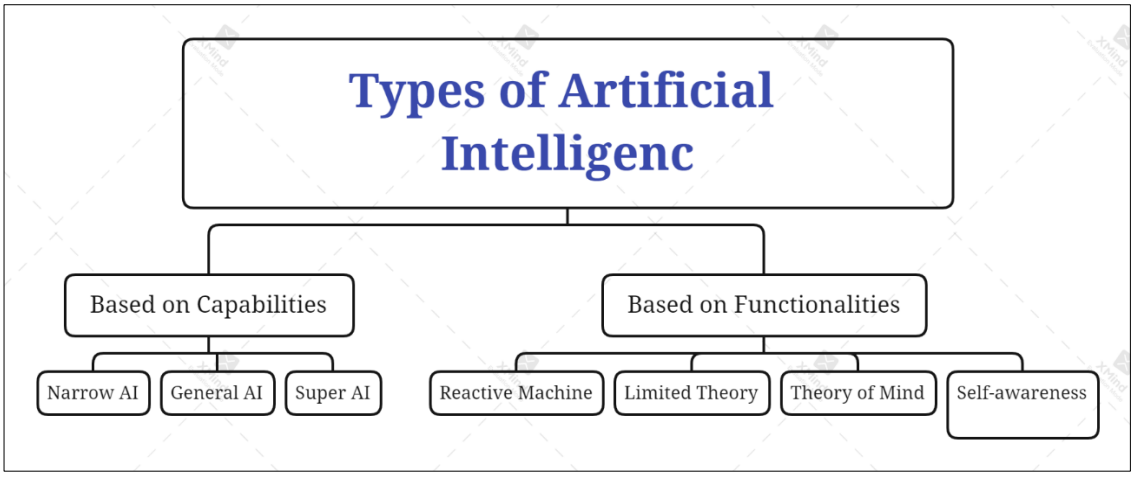
The process of creating intelligent machines out of massive amounts of data is known as artificial intelligence. Systems execute human-like tasks by learning from their prior knowledge and experiences.[75] It improves the efficiency, effectiveness, and speed of human endeavours. The process of creating intelligent machines out of massive amounts of data is known as artificial intelligence.[76]

Although there are many different ways to categorize artificial intelligence, there are primarily two basic groups that are based on the capabilities and functionalities of AI.

---

[73] Cp. *Christopher Collins et al.*, Information Systems Research, 2021, p. 102383.
[74] See https://www.proquest.com/openview/943bd867a26bf544f7c2212f81314ee0/1?pq-origsite=gscholar&cbl=35418, accessed 03.07.2022; *No Author*, 2022, The End Of The Human Era, no page number.
[75] Cp. *Patrick Brézillon.*, Context in Artificial Intelligence, no page number.
[76] Cp. *Daniel E. O'Leary.*, Artificial Intelligence and Big Data, 2013, pp. 96–99.

Figure 13: Representation of Types of Artificial Intelligence



Source: Author's own representation

AI type-1: Based on Capabilities

Narrow AI: A kind of AI known as narrow AI is capable of intelligently carrying out a certain task. In the field of artificial intelligence, narrow AI is the most prevalent and readily available AI. Since narrow AI is only taught for a single job, it cannot accomplish tasks outside of its domain or set of constraints.[77]

Given that it combines an Expert system method together with machine learning and natural language processing (NLP), IBM's Watson supercomputer also falls under the category of Narrow AI. Playing chess, receiving recommendations for purchases on an e-commerce site, self-driving automobiles, speech recognition, and image identification are certain instances of narrow AI.

General AI: An intelligence called general artificial intelligence (GAI) is capable of doing any intellectual job as effectively as a person. The goal of general AI is to create a system that is intelligent enough to think like a person on its own.[78] There is presently no system that falls under general AI that can carry out every work as effectively as a person.

Super AI: Super AI is a degree of system intelligence where computers are capable of outperforming people in any task owing to their cognitive abilities.[79] It results from

[77] Cp. *Osmo Kuusi and Sirkka Heinonen.*, Artificial General Intelligence, 2022, p. 194675672211016.
[78] Cp. *Ben Goertzel.,* Artificial General Intelligence, 2014, pp. 2013–15.
[79] Cp. *Yutong Lu et al.*, Super-AI Machines, 2018, pp. 82–87.

general AI. Strong AI has the capacity to understand, reason, solve puzzles, make judgments, plan, learn, and communicate on its own, among other crucial features.

AI type-2: Based on the functionality

Reactive Machines: Perhaps the most fundamental forms of artificial intelligence are machines that are only reactive. Such artificial intelligence systems don't retain memories or past experiences for future use. These machines just concentrate on the present situation and respond in the best way they can.[80]

Reactive machines include IBM's Deep Blue system. Another example of a reactive machine is AlphaGo from Google.

Limited Memory: Machines with a small amount of memory can temporarily store some data or memories of the past. Only a brief length of time can be used by these machines to access stored data.[81]

One of the best instances of Limited Memory systems is self-driving automobiles. These vehicles can retain information for road navigation, including the speed limit, distance to other vehicles, and recent speeds of those nearby.

Theory of Mind: Theory of Mind AI should have been able to connect socially with humans and comprehend human emotions, people, and opinions. Although this kind of AI machine has not yet been built, scientists are working hard to advance their capabilities.[82]

Self-Awareness: Self-awareness Artificial intelligence is where technology is headed. They will be extremely clever and possess consciousness, feelings, and self-awareness. These devices will be more intelligent than a human mind. Self-Awareness AI is still a theoretical idea and may not now actually exist.

2.4.4 Overview of Artificial Intelligence

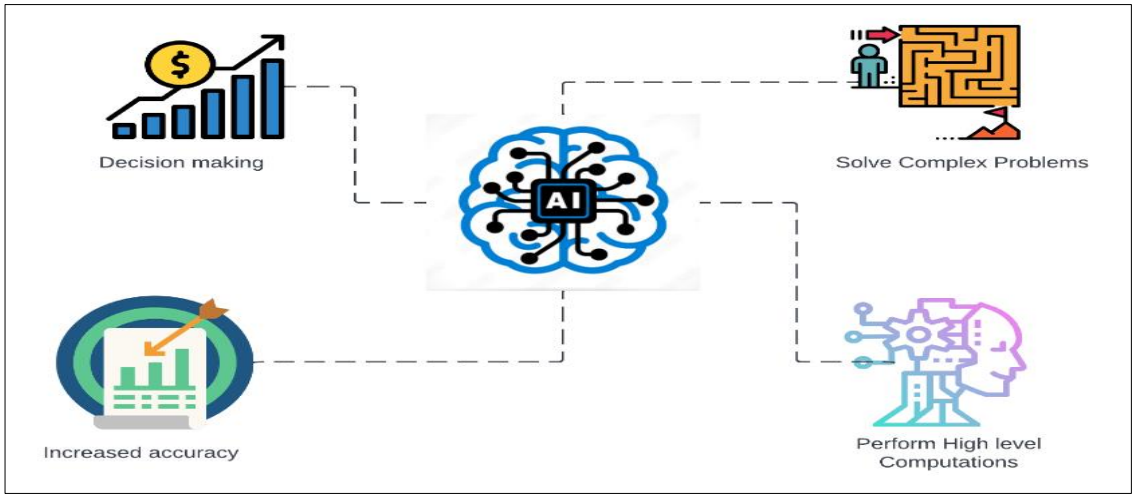
The development of intelligent machines has advanced significantly due to the interdisciplinary field of AI. The major objective of AI has been to create an intelligent and autonomous system that is capable of learning in many more dimensions and

[80] Cp. *Rajeev Alur and Radu Grosu.*, Modular Refinement, 2000, pp. 390–402.
[81] Cp. *Andrew H. Jazwinski*, Limited Memory Optimal Filtering, 1968, pp. 558–63.
[82] Cp. *Chris Frith and Uta Frith*, Theory of Mind, 2005, pp. 644–45.

evaluating data.[83] For evaluating vast amounts of data and spotting trends, machines are useful. If this is successful, it will be able to generate projections that are significantly more accurate and efficient than those made by individuals.

Despite the fact that AI has numerous fields, it will be difficult to account for all of them. Because the research is based on the insurance sector, the main goal of the next part is to clarify which aspects of technology are used within it.

Figure 14: Representation of Artificial Intelligence



Source: Author's own representation

Deep Learning:

Deep learning is based on the field of machine learning, which is a part of artificial intelligence. Deep learning will operate because neural networks mimic the functioning of the human brain.[84] Nothing is controlled directly in deep learning. In general, it is a class of machine learning that does feature extraction and transformation using a large number of nonlinear processing units. Each of the next layers uses the output from the one below as its input.

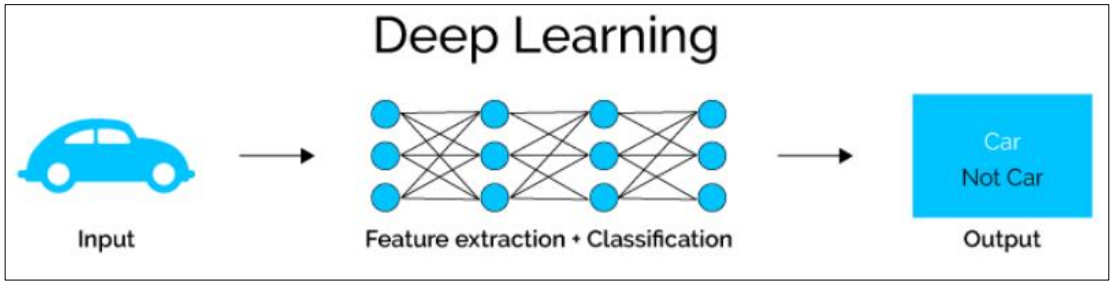
Specifically, when there are many inputs and outputs, deep learning methods are used. Although machine learning, a subset of artificial intelligence is where deep learning

[83] See http://www.scimagojr.com/shapeofscience/., accessed 10.06.2022; *Fernando Martínez-Plumed et al.*, 2022, The Facets of Artificial Intelligence, no page number.
[84] Cp. *Yann Lecun, Yoshua Bengio, and Geoffrey Hinton*., Deep Learning, 2015, pp. 436–44.

originated, and since the goal of artificial intelligence is to mimic human behaviour, so too is the goal of deep learning to construct such an algorithm that can mimic the brain.[85]

Neural networks are used to perform deep learning, and the brain cells that make up biological neurons serve as the inspiration for these networks. In fact, deep networks, which are nothing more than neural networks with several hidden layers, are used to perform deep learning.

Figure 15: Representation of Deep Learning



Source: Based on https://semiengineering.com/deep-learning-spreads/, accessed 03.07.2022

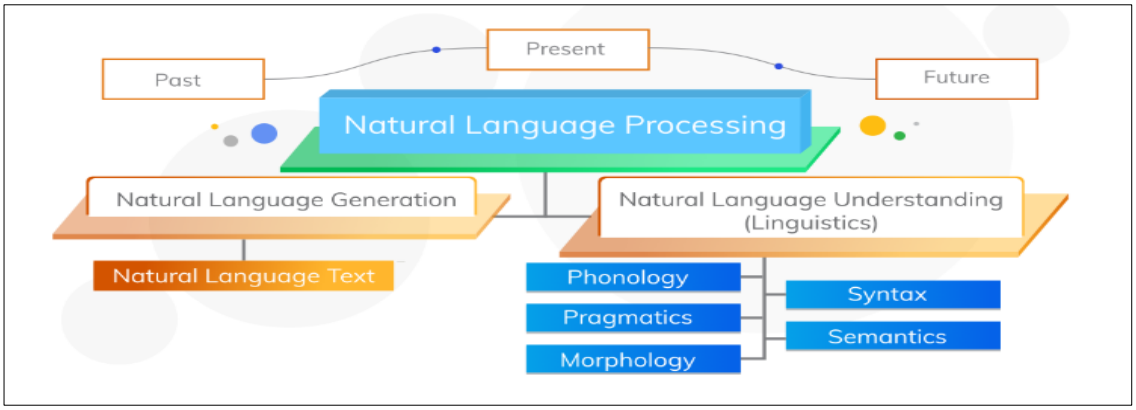
Natural Language Processing (NLP):

Natural Language Processing, or NLP for short, is a subfield of computer science, humanities, and artificial intelligence. Machines can understand, analyse, manipulate, and interpret human languages thanks to technology.[86] In order to execute tasks like translation, automatic summarization, Named Entity Recognition (NER), speech recognition, relationship extraction, and topic segmentation, it aids developers in organizing knowledge. Some of the most important methods that NLP is used are Text Summarization, Sentimental Analysis, machine translation, etc.

As an example, IBM has developed a tool to help with NLP operations and tasks. One of the promising textual analysis products on the market is IBM's Intelligent Miner. It offers several features, including text summarization, word clustering, text categorization, and fundamental text cleaning, all of which are necessary to get positive outcomes.

[85] Cp. *Yanming Guo et al.*, Visual Understanding, 2016, pp. 27–48.
[86] Cp. *Yue Kang et al.*, Natural Language Processing, 2020, pp. 139–72.

Figure 16: Representation of Natural Language Processing



Source: Based on https://www.xenonstack.com/blog/evolution-of-nlp, accessed 03.07.2022

Programming Languages used to build AI models:

The goal of AI, a machine learning technique, is to create models or programs that can execute tasks well or better than a human. AI is now regarded as a weakling as well. Due to the fact that it only does a specific task, like facial recognition. On the other hand, AI also uses algorithms, which are a collection of procedures for doing calculations and problem-solving operations.[87] Programming languages are used to solve complex algorithms, simplifying calculations and user interactions. The languages that are most commonly used for programming AI models are python, R programming, and so on.

2.5 The Business Model

The term plan of action suggests an organization's arrangement for creating a gain.[88] It recognizes the items or administrations the field-tested strategies to sell, its distinguished objective market, and any expected costs. Plans of action are significant for both new and set up organizations.

Throughout recent many years, the expression ‘plan of action’ has every now and again been abused by the two scholastics and experts.[89] It is normal to hear the term being utilized by administrators, specialists, or researchers from assorted fields and surprisingly in the famous media. After some time, the expression ‘plan of action’ has been

[87] Cp. *Onyeka Ezenwoye*, Programming Language, 2019, no page number.
[88] Cp. *Andrea Ovans*, Business Model, 2015, no page number.
[89] Cp. *Ramon Casadesus-Masanell and Joan E Ricart.*, A Winning Business Model, no page number.

experienced in two fundamental ways: first, it has developed into an indistinct thought with ripping apart propensity towards other administration terms, for example, "system"; and, second, a few organizations during the 1990s were directed to terrible showing and at last liquidation because of following what was probably imaginative business models. The time has come to relearn what the expression "plan of action" incorporates and demonstrate its significance and utility to both the scholarly and the business local area.

Today, plan of action prospects inside the media transmission industry are tremendous. New and creative methods of carrying on with work are being found at a quicker pace than at any other time.[90] In this way, propels in innovation permit cell phone producers to create incomes not just from the offer of their handsets and related embellishments yet in addition from a few different sources. As the negligible expenses of going through with exchanges in a computerized world are near nothing, cell phones have become a billion-dollar dispersion station where a large number of computerized items like music, motion pictures, photographs, programming, and games are bought and consumed immediately.

A typical error many organizations make, when they make their plans of action is to underrate the expenses of financing the business until it becomes productive. Counting expenses for the presentation of an item isn't sufficient. An organization needs to keep the business running until its income surpasses its costs.

At the point when endeavours contend with utilizing plans of action that vary from each other, the results are challenging to foresee.

One plan of action might seem better than others when dissected in confinement however make less worth than the others when communications are thought of. Or on the other hand, adversaries might wind up becoming accomplices in esteem creation. Evaluating models in an independent style prompts flawed appraisals of their assets and shortcomings and terrible navigation. This is a central motivation behind why so many new plans of action fizzle. Furthermore, the penchant to overlook the unique components of plans of action brings about many organizations neglecting to utilize them to their maximum capacity.

[90] Cp. *Eelis Rytkönen and Suvi Nenonen*., The Business Model Canvas, 2014, pp. 138–54.

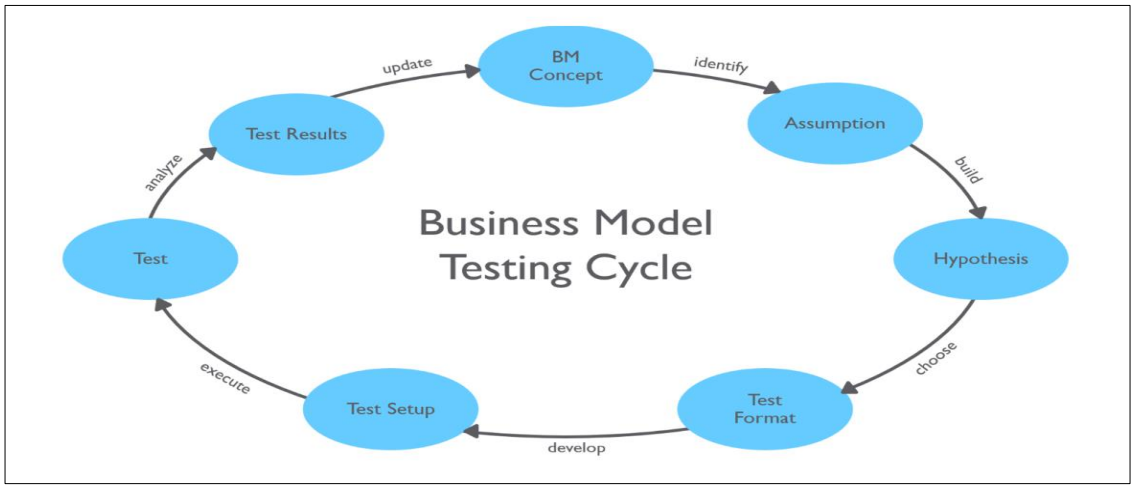
There are as many sorts of business models as there are kinds of business. For example, direct deals, diversifying, promoting based, and physical stores are on the whole instances of conventional plans of action. There are hybrid models too, for example, organizations that consolidate web retail with physical stores or with donning associations like the NBA.

Importance of Business Model:

In the corporate world, the relationship between company design, strategy, and innovation are connected irreversibly. Business modelling, as a business design process, is a strategic activity that encourages innovation as part of a larger business plan.[91] The business model has forever been a significant piece of carrying on with work, yet as of late the planning of organizations got another direness. It is the development of new companies, rapidly evolving markets, and advancing mechanical potential outcomes that make it always critical to foster creative plans of action and reconsider conventional ones. These conditions abbreviate the time that old plans of action can remain feasible.

Business Design gives the apparatuses and techniques to create and test a plan of action.[92] Simultaneously it has the effect between assistance that is only a promoting trick and one that is a supportable business.

Figure 17: Representation of Business Model Testing Cycle



[91] Cp. *Henry Chesbrough*, Business Model Innovation, 2010, pp. 354–63.
[92] See https://books.google.de/books?hl=en&lr=&id=0X54fSKq7bkC&oi=fnd&pg=PR13&dq=business+design+techniques&ots=54s9fDMpQ1&sig=cQrJrYNngWfAAXMtn3BUKej-j-w&redir_esc=y#v=onepage&q=business design techniques&f=false, accessed 22.06.2022; *K.N.·奥托 (美), Kevin N. Otto.*, 2022, Reverse Engineering and New Product Development, no page number.

Source: Based on https://bmilab.com/blog/2019/4/18/the-big-challenge-in-business-model-development-how-to-systematically-test-your-business-model, accessed 25.06.2022

At the Board of Innovation, we have a comprehensive way to deal with business demonstrating. We plan organizations the same way we plan items. Joining the objective section and client research, we complete extensive market and contender research. We centre around laying out the predominant plans of action and most significant long-held convictions that can be disturbed. Whenever we model arrangements, we plan different pertinent plans of action along with you. Our item surveys are trailed by the plan of action assessment works out. Lastly, while evaluating various models, we explore different avenues regarding different business models, as well. Our 3-day business model development run is organized into 4 phases which are detailed underneath.

1) Business model research:

The main purpose is to understand the market, identify unknowns that need to be further researched,

- Recognize the predominant plans of action and most significant long-held convictions that can be disturbed.
- Get the numbers, KPIs, and benchmarks behind your industry, market, and business.
- Recognize the main patterns and how they will change your industry.
- Recognize the significant things you have hardly any insight into the business, that you should ask during interviews with partners.

2) Business model ideation:

3-5 distinct business models that could work for the business,

- Make an interpretation of long-held convictions to potentially open doors for disturbance.
- Distinguish plans of action that you could work (Business Model flowchart).
- Ideate your plan of action highlights with ideation practices made to your undertaking.

3) Modelling your business:

Structure recently ideated plans of action and uncover hidden suppositions.

- Model your recently ideated plan of action with esteem.

- Trades (utilizing the plan of action unit) to make it concrete.
- Sum up the main parts of your plan of action and recognize hidden suppositions.

4) Evaluation of the business models:

figure out if the business model a person came up with has a chance of survival in the real world

- Fabricate your business industrial facility utilizing development measurements and distinguish your possibly greatest difficulties under a working plan of action that you can begin testing immediately.
- Gauge the cost of acquisition (COA) and lifetime value (LTV) of your clients. Figure out what will this mean for your business.
- Fabricate your scale model and comprehend your forthcoming scaling difficulties early so that you can begin testing right away.

In light of your model, what are the models and extraordinary elements of all-around planned plans of action, go-to-showcase methodologies, scaling strategies, and things you should remember while pushing ahead.

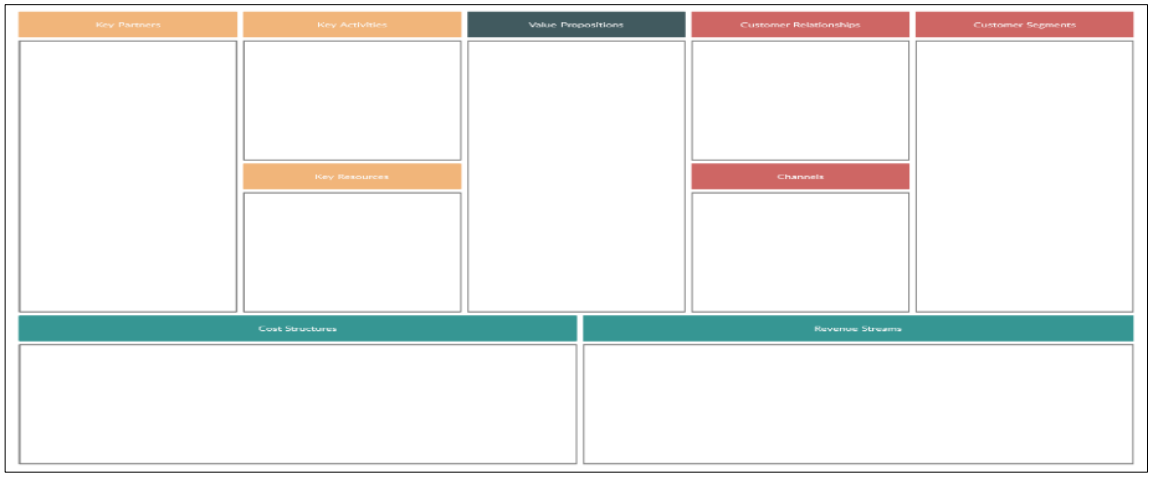
2.6 The Business Model Canvas

The Business Model Canvas (BMC) is a visual modelling method that is used to capture the business model of a company.[93] It is defined by nine building blocks. The Business Model Canvas (BMC) is an essential administration apparatus to rapidly and effectively characterize and impart a business thought or idea. It is a one-page record that deals with the basic components of a business or item, organizing thought in a rational way. The right half of the BMC centres around the client outside, while, the left half of the material spotlights the business interior. Both outside and interior variables meet around the offer, which is the trading of significant worth between your business and your client/customers. The business model canvas beats the conventional marketable strategy that ranges across a few pages, by offering a lot more straightforward method for getting the different centre components of a business.

[93] Cf. *Alexander Osterwalder, Yves Pigneur, and Christopher L. Tucci.,* Clarifying Business Models, 2005, p. 1.

The right half of the material spotlights on the client or the market (external factors that are not influenced quite a bit by) while the left half of the material spotlights on the business (internal factors that are for the most part under influence).[94] In the centre, it gets the offers that address the trading of significant worth between your business and your clients. The Business Model Canvas is upheld by four fundamental areas of business: Offering, Customers, Infrastructure, and Financials. These regions incorporate out and out nine structure blocks: Offering comprises of the value proposition; customer relationships, channels, and customer segments compose Customers; Financial includes cost construction and income streams; and key partners, key activities, and key resources establish Infrastructure.

Figure 18: Representation of Business Model Canvas



Source: Author's own representation

1. Customer Segments

The Customer Segments Building Block defines the various gatherings or associations a venture intends to reach and serve. Clients contain the core of any plan of action. Without clients, no organization can get by for long.[95] To all the more likely to fulfil clients, an organization might bunch them into unmistakable portions with normal requirements, normal practices, or different properties. A plan of action might characterize one or a few enormous or little Customer Segments. An association should settle on a cognizant choice

[94] See https://medium.com/seed-digital/how-to-business-model-canvas-explained-ad3676b6fe4a, accessed 24.06.2022; *Sheda, Sheda.*, Business Model Canvas Explained, no page number.

[95] Cp. *Christopher P. Blocker and Daniel J. Flint.*, Customer Segments, 2007, pp. 810–22.

with regards to which sections to serve and which portions to disregard. Whenever this decision is taken, a strategy may be methodically devised based on a thorough understanding of the client's stated requirements.[96]

Client groups address separate fragments if:

- Their necessities require and legitimize an unmistakable proposition.
- They are reached through various Distribution Channels.
- They have substantially different profitability.
- They will pay for various parts of the deal.

2. Value propositions

The Value Propositions section portrays the heap of items and administrations that provide value to a certain Customer Segment.[97] The Value Proposition is the motivation behind why clients go to one organization over another. It tackles a client's concern or fulfils a client's need. Each Value Proposition comprises a chosen heap of items or potential benefits that obliges the necessities of a particular Customer Segment. In this sense, the Value Proposition is a conglomeration, or group, of advantages that an organization offers clients.[98] A few Value Propositions might be creative and address a new or problematic offer. Others might be like existing business sector offers, however with added highlights and traits.

3. Channels

The Channels Building Block depicts how an organization speaks with and arrives at its Customer Segments to convey a Value Proposition Communication, appropriation, and deals Channels include an organization's connection point with clients.[99] Channels are client contact focuses that assumes a significant part in the client experience.

---

[96] Cp. *Sally Dibb and Lyndon Simkin.*, Quality of Customer Segments, 2010, pp. 113–31.

[97] See https://books.google.de/books?hl=en&lr=&id=iemTssZgYlwC&oi=fnd&pg=PA21&dq=value+proposition+&ots=hp_n_t2_S4&sig=wS_pL9ZuytC0PZhVjycfJ9bG7W8&redir_esc=y#v=onepage&q=value proposition&f=false, accessed 22.06.2022; *David Ulrich, Wayne Brockbank.*, Value Proposition, no page number.

[98] Cp. *Adrian Payne, Pennie Frow, and Andreas Eggert.*, The Customer Value Proposition: Evolution, no page number.

[99] See https://bmcintroduction.wordpress.com/channels/, accessed 25.06.2022; *No Author*, Business Model Canvas, no page number.

Channels serve a few capacities, including:

- Bringing issues to light among clients about an organization's items and administrations.
- Assisting clients with assessing an organization's Value Proposition.
- Permitting clients to buy explicit items and administrations.
- Giving post-buy client care.

4. Customer Relationships

The Customer Relationships Building Block depicts the kinds of connections an organization sets up with specific Customer Segments.[100] An organization ought to explain the kind of relationship it needs to set up with every Customer Segment. Connections can go from individual to computerized.

Customer connections might be driven by the accompanying inspirations:

- Customer obtaining
- Customer maintenance
- Supporting deals

The Customer Relationships called for by an organization's plan of action profoundly influence the general client experience.

5. Revenue Streams

The Revenue Streams sector addresses the money an organization produces from every Customer Segment. Assuming that clients include the core of a plan of action, Revenue Streams are its supply routes. An organization should ask itself, for what worth is every Customer Segment really ready to pay? Effectively addressing that question permits the firm to produce at least one Revenue Streams from every Customer Segment.[101] Every Revenue Stream might have different evaluating instruments, for example, fixed rundown costs, haggling, unloading, market subordinate, volume ward, or yield the executives.

6. Key Resources

---

[100] Cp. *Michelle C Burger Mangindaan and Evelyn Hendriana.*, Customer Relationship Management, 2013, no page number.

[101] Cp. *Christoph Zott, Raphael Amit, and Lorenzo Massa.*, The Business Model, 2011, pp. 1019–42.

The Key Resources section portrays the main resources needed to make a plan of action work. Each plan of action requires Key Resources. These assets permit an undertaking to make and offer a Value Proposition, arrive at business sectors, keep up with associations with Customer Segments, and procure incomes.[102] Different Key Resources are required relying upon the kind of plan of action. A central processor producer requires capital-escalated creation offices, while a CPU architect zero in additional on HR. Distinct advantages can be physical, monetary, scholarly, or human. Secret weapons can be claimed or rented by the organization or procured from key accomplices.

7. Key Activities

The Key Activities section portrays the main things an organization should do to make its plan of action work. Each plan of action requires various Key Activities.[103] These are the main moves an organization should make to work effectively. Like Key Resources, they are needed to make and offer a Value Proposition, arrive at business sectors, keep up with Customer Relationships, and acquire incomes. Furthermore, like Key Resources, Key Activities contrast contingent upon plan of action type.[104] For programming producer Microsoft, Key Activities incorporate programming improvement**.**

8. Key Partnerships

The Key Partnerships section portrays the organization of providers and accomplices that make the plan of action work.[105] Organizations fashion associations for some reasons and associations are turning into a foundation of numerous plans of action. Organizations make coalitions to advance their plans of action, lessen hazards, or gain assets.

It can recognize four distinct sorts of organizations:

- Vital partnerships between non-contenders.
- Coopetition: vital organizations between contenders.
- Joint dares to foster new organizations.

---

[102] Cp. *Michael E Porter et al.,* Reinventing Your Business Model, 2000, no page number.

[103] Cp. *Daniel Kindström and Mikael Ottosson.,* Energy Services, 2016, pp. 491–500.

[104] Cp. *Wiebke Reim et al.,* Implementing Sustainable Product–Service Systems, 2017, pp. 61–66.

[105] See https://www.researchgate.net/publication/291300642, accessed 25.06.2022; *Ronald W Mcquaid,* The Theory of Partnership, no page number.

- Purchaser provider connections to guarantee solid supplies.

9. Cost Structure

The Cost Structure portrays all expenses brought about to work a plan of action. This building block portrays the main expenses brought about while working under a specific plan of action.[106] Making and conveying esteem, keeping up with Customer Relationships, and producing income all-cause costs. Such expenses can be determined moderately effectively in the wake of characterizing Key Resources, Key Activities, and Key Partnerships. Some plans of action, however, are more expense-driven than others. Purported "nitty-gritty" carriers, for example, have assembled plans of action completely around minimal expense Structures.

2.7 Automobile Industry

The automobile industry has its beginnings in the late 19th century when businessmen from all over the world attempted to carve out a position for themselves in this brand-new and prosperous society.[107] Nevertheless, one of the key issues that this just formed sector had to cope with had already been present. In the 19th century, a car was a locomotive that had been downsized. It would have wheels attached to its own chassis and a steam engine. Because steam engines were thought to be too dangerous because they couldn't be utilized with earlier materials for automobiles, the government didn't wait long to pass new laws to stop such cars from being on the roads.[108] And this marks the beginning of the history of the automobile as people know it today. The first person to patent a suitable thermic explosion engine for the road was Mr. Daimler, who is also the true creator of Mercedes-Benz. Ever since automobile manufacturers have engaged in an innovation competition.

The invention is a necessity and it drastically changes the market. To exemplify, Volvo holds the patent for all safety belts used in modern automobiles. This perspective allows one to declare that the automobile business has always been a tough playground of ruthless competition, invention, and sharing between automobile makers.

---

[106] Cp. *Daniel M Russell et al.*, The Cost Structure of Sensemaking, no page number.

[107] Cp. *R. J. Orsato and P. Wells.*, The Automobile Industry, 2007, pp. 989–93.

[108] See https://www.jstor.org/stable/20081778, accessed 26.06.2022; History of the Automobile, July 3, no page number.

Nowadays, cars are getting ‘smarter’, more ‘connected’, and more ‘conscious’.[109] It is now possible for vehicles to sense objects, make autonomous decisions, and direct you from point A to point B while taking into account the demands and preferences of the driver. Both technical revolutions directly led to the development of automobiles.[110] They were developed during the Industrial Revolution, and after the advent of the digital revolution in the 1990s, their sole function underwent a full redefinition, which has persisted until the present day. As one can see, the car market adjusts in line with how the globe is changing.

2.7.1 Definition of Automobile Industry

The automobile has its meaning that is split into two-word auto (self) and mobile (movable). Therefore, the automobile is a self-movable vehicle used for transportation. The modern automobile is a sophisticated technical system that uses various subsystems, each of which performs a particular design role. Some of them have thousands of individual parts and were made possible by new or improved technologies, such as electronic computers, high-strength polymers, and novel steel and nonferrous metal alloys, or by advancements in already existing technologies. Various subsystems have developed as a result of things like air pollution, safety regulations, and global manufacturer competitiveness.

The automobile industry is the manufacturing and sale of self-powered vehicles such as passenger cars, trucks, farm equipment, and other commercial vehicles. The auto industry has promoted the development of an extensive road network, makes it easier for the progress of suburban areas and shopping centres around major cities, and performed a crucial role in the growth of related industries, such as the oil and travel industries, by allowing consumers to commute large distances for employment, shopping, and entertainment.[111] One of the biggest consumers of several essential industrial items, including steel, is now the auto sector. The industry's size has made it a significant factor in determining economic growth.

---

[109] Cp. *Melanie Ooi.*, Future Trend in IM, 2019, pp. 33–34.
[110] Cp. *R. J. Orsato and P. Wells.*, The Rise and Demise of the Automobile Industry, 2007, pp. 994–1006.
[111] Cp. *Anoop C. Nair, S. Ramalingam, and Ashvini Ravi.*, Knowledge Creation, 2015, no page number.

2.7.2 Role of Automobile Industry

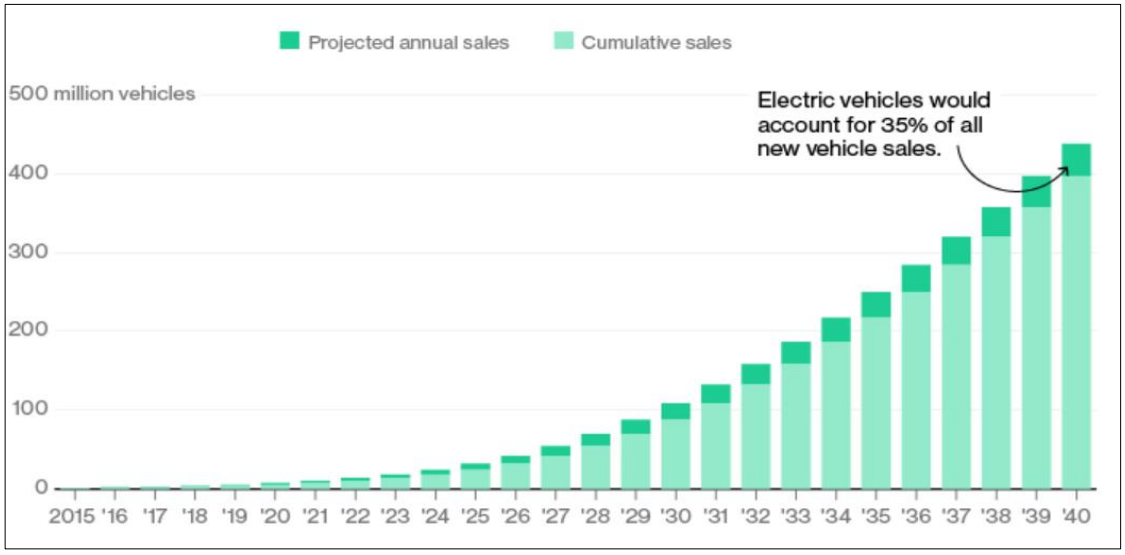
The production of environmental-friendly vehicles by automakers has been a trend for more than ten years.[112] The definition includes hybrid vehicles (half electric, half gasoline, like the Toyota Prius) and entirely electric vehicles, like Tesla cars, in addition to vehicles with 'eco' modes (like the Honda Civic), which control fuel usage by altering gear shifting up. Electric vehicles have existed for some time.[113] They were invented about 120 years ago, but the efficient gas-powered engine never overcame them since electric technology at the time lacked both power and range.

However, statistics and public perceptions have recently shown that electric cars are back and supported by new technology that makes them more dependable, effective, and affordable. It's interesting to emphasize that there are other elements to consider for the success of electric automobiles in addition to technological ones.

Additionally, it is a result of a variety of environmental rules, such as the first national regulation of greenhouse emissions under Obama. Producing electric vehicles also gives the auto industry a chance to greenwash its image and promote itself as an environmentally conscious business. Numerous brands use this marketing technique, whether they are in the automobile business or another one like the tech sector or retail.[114] Businesses that use 'eco-friendly tools' help customers feel like members of a community that values the environment and makes them aware of their place in it.

[112] Cp. *Mario Cools et al.*, Environment-Friendly Modes, 2009, pp. 437–53.
[113] Cp. *Ajitha and Nagra.*, Artificial Intelligence in Automobile Industry, no page number.
[114] Cp. *Taizo Masuda et al.*, Automobile Energy Source, 2016, pp. 580–84.

Figure 19: Representation of Rising of Electric Vehicle sales



Source: Based on http://www.behindenergy.com/how-electric-cars-will-cause-the-next-oil-crisis/?lang=en, accessed 05.07.2022

There is little doubt that electricity will play a significant role in the future, with more rivals and cheaper than ever cars. The large luxury brands, including Ferrari, Lamborghini, Maserati, and McLaren, represent a niche market that will either not be affected at all or just somewhat by this trend. With their powerfully revving engines, such brands have achieved fame.

The automobile industry is constantly making technological and mechanical advancements. An automobile manufacturer, like Lamborghini, develops every innovation to set itself apart from Ferrari, which is its biggest rival. They are also applied to meet a specific requirement in their clients' expectations, resulting in superior customer experience behaviour.

A recent incident occurred in 2015 when Volkswagen, a car manufacturer, was found to have cheated by employing software to pass particle pollution regulations on their vehicles. Due to this significant scandal, VW suffered a severe loss of profit. Market shares dropped by 20% compared to before the controversy. The VW group also reported a 1.87-billion-dollar loss as a result of having to set money aside to pay a fine that would have cost them 22 billion dollars in retrieving and fixing all the defective vehicles in order to make up for the harm they had caused.

In addition to having to recall all the cars and remove the cheating software, they also had to ensure that the restored cars legitimately passed the emission tests. Volkswagen used the customer behaviour tool very effectively because as soon as the issue broke, the CEO personally apologized and dissociated the scandal from VW's guiding principles.

As a result, the company paid attention to their clients' psychological needs, and because the CEO addressed them directly, they gave the appearance of a personal relationship. When VW announced comprehensive cooperation between the business and the government in addition to customer care, the social element of the crisis was entirely covered. Volkswagen's quick action reduced the losses and resulted in a highly effective comeback over the course of a reliable campaign.

Consequently, the auto industry needs to take ecological factors into account. The Volkswagen example demonstrates how norms, laws, and regulations can weaken this sector of the economy and how opposing ideas can exist.

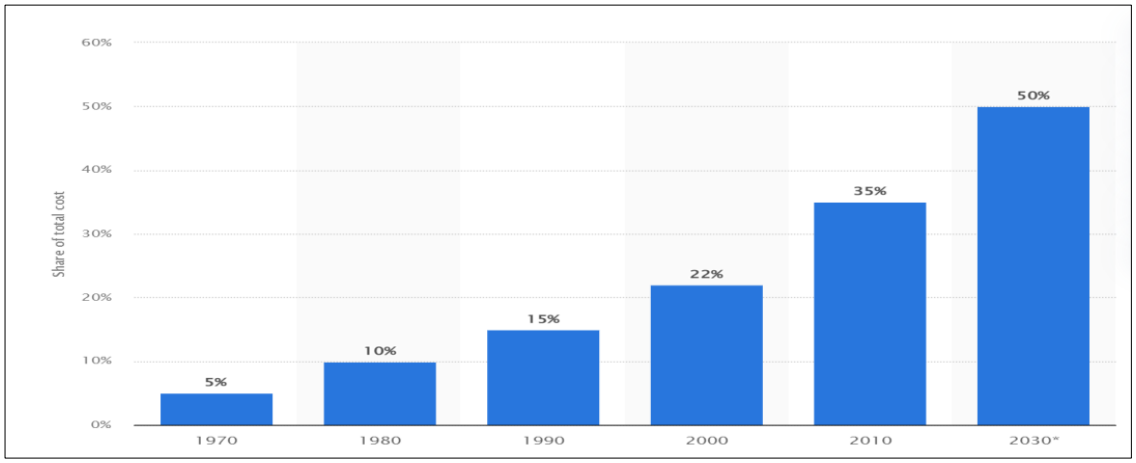
Since they are no longer made to just move people from one place to another, cars have evolved into a means of mobility as well as a means of transportation. The goal of automakers is to do this by continuously implementing new technology. In the past ten years, technology has advanced incredibly and taken over not just our daily lives (at home, at work, etc.), but also our cars.

In a disruptive way, the digitalization of the automobile industry is a game-changer. It drives innovation and creativity and alters the guidelines set forth by previous business models, paving the way for more risk-taking ones that, depending on the model picked, also present greater value creation or co-creation as well as value capture.[115] Start-ups are places where innovative ideas and business concepts can be highly valued; one such example is Uber, the most valuable start-up in the world, with a valuation of $60 billion. The auto industry receives a significant boost as a result, leading the "existing player" (also known as the major producers Volkswagen and Toyota) to adapt and study their more innovative competitors. As a result, BMW released its app BMW drive today, and VW is just getting started. By 2030, the automobile industry is predicted to have adopted a total of 6 significant technical advancements. Connected objects, shared connections,

[115] Cp. *Philippe Aghion, Céline Antonin, and Simon Bunel.*, Artificial Intelligence, 2019, pp. 149–64.

data, robotics, anticipation, and a general sense of self-awareness from the car are all among them.

Figure 20: Representation of Automobile electronics cost as a percentage of total car cost worldwide from 1950 to 2030



Source: Based on Derenda et al., Road transport, 2018, p. 4

The first company to ever introduce a self-driving automobile was Google. Its rooftop was covered with numerous cameras and using software, it could use the images and sensors (radars) to travel in a secure manner (without hurting people). Although Google was able to pioneer this development in automotive technology, they haven't yet commercialized a single-vehicle. Another firm, created by Elon Musk and called Tesla, has quickly embraced this trend of autonomous vehicles and made it one of its primary value propositions in addition to having totally electric vehicles (100 percent electric). The traditional driving aids that automakers have been promoting up until now, such as lane assist (which beeps when one crosses the white line of the road), sleep assist (which beeps when it detects a sleepy attitude in the user's behaviour), and dead angle assists (which causes the side mirrors to blink when a car passes in the dead angle of the user's car), could be seen as evolving into autonomous driving.[116] All of these technological advancements are combined into the autonomous driving experience, and satnav (satellite navigation) assistance is added on top of it. The vehicle can perceive its surroundings and can avoid obstructions or stop in response to the automobile in front thanks to the radar sensors distributed throughout the vehicle. However, the benefits of the self-driving experience are not as significant because it is only available in a small

[116] Cp. *Asif Faisal et al.*, Understanding Autonomous Vehicles, 2019, pp. 45–72.

number of automobiles (fewer accidents, more time to focus on other projects). Even though it hasn't even been introduced to the automotive industry, this technology already has version 2.0, which entails using certain signals to connect automobiles to one another, enabling them to communicate and make better judgments based on the data. Since the car would be operated robotically with no human involvement, there would be little margin for error, and increasing the insurance company's trust in the system would be beneficial.

Considering both digitalization and environmental factors, one may claim that the integrity of the vehicle sector is destined to change its structure. That suggests a change in its economy, and as a result, a change in the suggested payment formulas to the clients.[117] Market participants become better educated, more self-aware, and generally more informed. Customers will therefore have complete knowledge of what to buy, how much it will cost, and where to buy it in the future as a result of those aspects.

[117] Cp. *M. F. Ward.*, European Motor Car Industry, 2007, pp. 443–53.

3. Research Design

In the last two decades, research methods have grown to the point where investigators and inquirers have a plethora of options. It urges that individuals creating a proposal or plan use a general framework to provide guidance on all aspects of the study, from evaluating the fundamental philosophical ideas under investigation to the detailed data gathering and analysis techniques.[118]

3.1 Structured / Systematic Literature Review

A Structured Literature Review is an integral part of any research, and it involves sourcing, reviewing, and analysing previous scientific work to draw further conclusions. A structured literature review can be defined as a scientific effort to identify all empirical evidence that satisfies the pre-defined scope criteria to answer a particular research question or hypothesis defined in the study.[119]

Figure 21: Representation of systematic literature review

Source: Author's own representation

A literature review is an essential part of the research activity. Scientific progress must be based on the previous effort, and a thorough understanding of the current body of work can be gained by evaluating relevant literature and identifying research gaps that may need to be investigated. It can be tested particular hypothesis and/or establish new hypotheses by summarizing, evaluating, and synthesizing a collection of relevant material.[120] Alternatively, it may also use a criterion to assess the validity and quality of previous work, revealing flaws, inconsistencies, and contradictions.

---

[118] Cp. *No Author*, Encyclopedia of Research Design, 2022, no. page number.
[119] Cp. *No Author*, Strict Structured Literature Reviews, 2018, no. page number.
[120] Cp. *Yu Xiao and Maria Watson.*, Conducting a Systematic Literature Review, 2019, pp. 93–112.

Literature reviews, like scientific investigations, should be legitimate, dependable, and reproducible. A systematic literature review (SLR) identifies, selects, and critically evaluates research material to address a specific topic.[121] The systematic review should adhere to a well-defined procedure or strategy, with the criteria clearly established prior to the start of the process. It's a transparent, thorough search that spans several databases and grey literature and can be duplicated by other academics. It entails devising a well-thought-out search strategy with a narrow emphasis on responding to a specific topic. Within established periods, the review indicates the sort of material searched, criticized, and reported. The evaluation must contain the search keywords, search tactics (including database names, platforms, and search dates), and restrictions.

Selection of relevant literature:

The relevant literature for the study was collected from various sources across the internet. A specific keyword search was conducted on platforms such as Google Scholar, Research Gate, Springer, Wiley, and various blogs to source the relevant literature. The gathered literature was further shortlisted based on a given set of criteria.

There are criteria for shortlisting the collected literature which are as follow:

1. Relevant keywords that were used in the study:
    - Automobile Industry
    - Traffic Congestion
    - Drive Assistance
    - Business Model Canvas
    - Artificial Intelligence
    - Machine Learning
    - Natural Language Processing
    - Structured Literature Review
    - Quantitative Data Analysis
    - Qualitative Data Analysis
    - Significance of Literature Review
    - Types of Research

[121] Cp. *Hannah Snyder*, Research Methodology, 2019, pp. 333–39.

2. Selection criteria:

Criteria that were used to select articles for the literature review based on research question. These are frequently referred to as inclusion and exclusion criteria. The requirements that an article must satisfy in order to be considered for inclusion in a literature review are known as inclusion criteria. It comprised research that compared various experimental therapies and was published within the previous five years.[122]

The second criteria which exclusion criteria was characteristics of a study that prevent it from being included in a literature review. An observational research design was adopted for the study.[123] Furthermore, qualitative methodology is used, research was released more than five years ago, and the study was published in a language other than English.

For this study, inclusion and exclusion criteria need to be balanced in order to generalize the study's findings.[124] Only a small group can be targeted if inclusion criteria are too rigorous and exclusion criteria too tight. Tight inclusion criteria frequently increase a study's internal validity, yet having a lot of exclusion criteria might significantly reduce the study's external validity. It is essential to strike a balance between the two to guarantee sufficient external validity.

There are several questions which posed when selecting and reviewing for this study:

- Do you have a clear set of research objectives before I start the review?
- Have you accurately cited every source that will aid in the formulation of my issue statement or research question?
- Have you analysed every type of literature, including publications on both qualitative and quantitative research?
- Do you have adequate theoretical or empirical evidence to back up my claim?
- Have you obtained enough proof from the literature to support the validity and accuracy of the designs or procedures I intend to utilize in my experiments?
- Have you determined the objective of the articles chosen for the literature review?
- Does your literature review be included in a report that has been written following a thorough investigation of the literature?

[122] Cp. *Lynne M. Connelly*, Inclusion and Exclusion Criteria, 2020, p. 125.
[123] Cp. *Ibid.*
[124] Cp. *Ibid.*

3.2 Research Methodology

All academic researches depend on the methodology and the clarity of reporting. Due to this the guidelines are to be specified on how and what is to be done and the set of rules that is to be followed.

3.2.1 Quantitative Method

Quantitative as the name suggests talks about quantity specially the numbers. This method usually deals with numbers and statistics. This method usually deals with finding relationships between variables and how they change with respect to each other (i.e., their correlations).[125] Quantitative research method is broadly utilized in the regular and sociologies: chemistry, psychology, chemistry, financial aspects, marketing, sociology, and so forth. Statistics, utilized in quantitative research, are a significant area of arithmetic and is broadly utilized when: (a) there is a need to dissect and deal with huge volumes of quantitative data to confirm speculations and to test a hypothesis, (b) there is vulnerability related to speculations viable, (c) exploration may be completed with surveys containing straightforward inquiries and short responses and (d) the information acquired can be measured and thought about. In quantitative research, data handling is commonly performed utilizing particular factual programming.[126]

There are quantitative research strategies for distinct, correlational, or trial research.

- In descriptive research, essentially look for a general synopsis of your review factors.
- In correlational research, people examine connections between your review factors.
- Experimental research is deliberately inspected whether there are circumstances and logical results connection between factors.

Quantitative research is frequently used to normalize information assortment and sum up discoveries.[127] Qualities of this approach include: Essential attributes of quantitative research approaches connected with explicit exploration targets are among others: (i) the

[125] Cp. *Suphat Kamolson.*, Quantitative Research, 2007, p. 20.
[126] Cp. *No Author.*, Quantitative and Qualitative Research, 2022, no page number.
[127] Cp. *Omar A. Ponce, et al.*, 3 Philosophy-Based Research, 2021, pp.35-52.

relationship of exploration with tests, (ii) the examination of peculiarities; (iii) the utilization of cutting edge apparatuses of measurements, (iv) the utilization of polls, (v) the evaluation of relations and highlights and (vi) the assortment, parade and show of quantitative data.

3.2.2 Qualitative Method

Qualitative research is an exceptionally expansive term that accepts research procedures that manage peculiarities by investigating encounters, practices, and relations without the utilization of insights and arithmetic and the handling of mathematical information.[128] Qualitative research is utilized to see how individuals experience the world.

There are several methods to qualitative research, they are:

- Observations: recording what exactly have seen, heard, or experienced indefinite field notes.
- Interviews: specifically asking individuals inquiries in one-on-one discussions.
- Focus groups: posing inquiries and creating conversation among a gathering.
- Surveys: dispersing polls with open-ended inquiries.
- Secondary research: gathering existing information as texts, pictures, sound or video accounts, and so on.

3.3.3 Mixed Methods

This is the most effective and most commonly used methodology in research. This method comprises of a combination of both qualitative and quantitative research methodology.

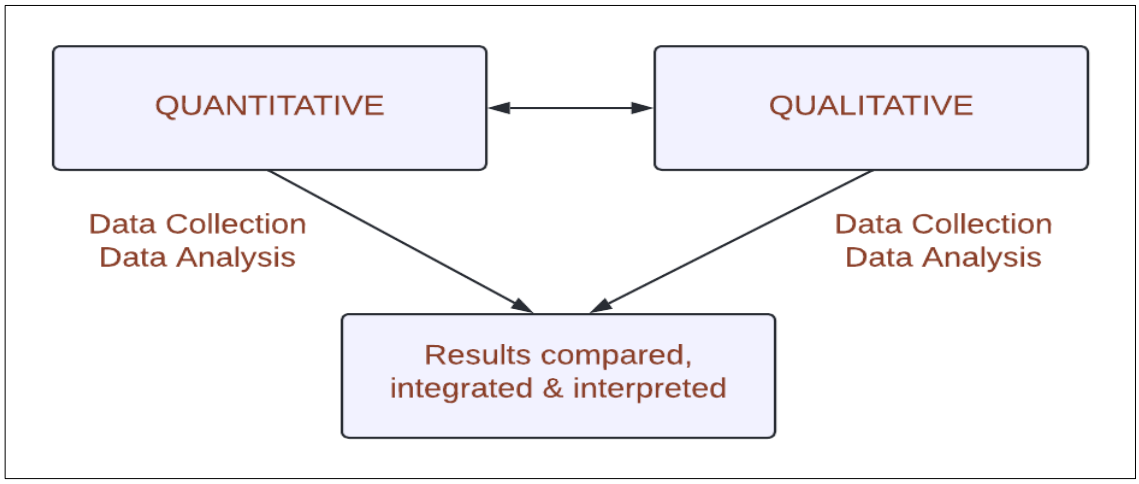
Some of the benefits of using mixed methods are:

- It offsets the disadvantages of the individual research methods and provides robust research results.
- It provides better interpretation and understanding of the data.

[128] Cp. *Loraine Busetto, Wolfgang Wick, and Christoph Gumbinger.*, Qualitative Research Methods, 2020, no page number.

This method uses mixed method of research as we do hypothesis testing and visualization of data to draw inferences with the case study of credit card fraud detection using predictive models.

Figure 22: Representation of mixed method



Source: Based on Prince Edem Dzakpasu, D., Hypothesis, 2017, p. 12

3.3 Justification of Research Approach

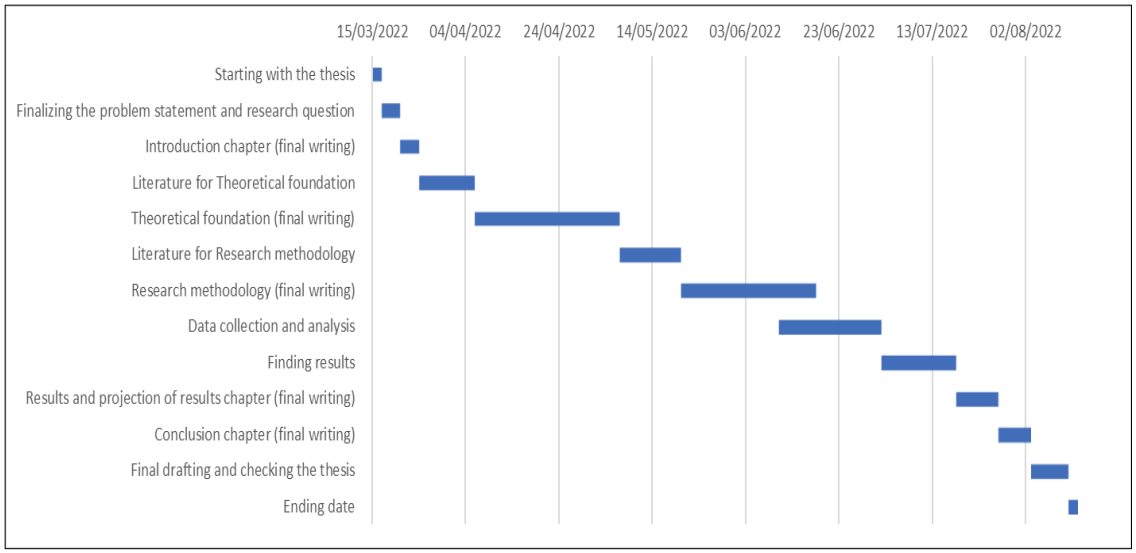
The study entitles quantitative research methodology of the state of Machine Learning and Artificial Intelligence in the automobile industry. It also uses the concepts of exploratory data analysis and detection by predictive modelling using machine learning algorithms. For the Literature review, different papers and intellectually informative sites were used and the papers were chosen in a way that they had a higher threshold of H-index and more recent works were chosen (2010-2021) with a few exceptions. For the case point of giving assistance to Drivers using predictive models were used by suggesting the shortest or best possible route for drivers with a sample data set procured from Kaggle.

3.4 Case point of giving assistance to Drivers using predictive models

This section explains one of the project approach and practical examples of adaption of Machine Learning and Artificial Intelligence through shortest possible route using a sample data set obtained from Kaggle.

3.4.1 Project plan

Figure 23: Representation of project plan



Source: Author's own representation

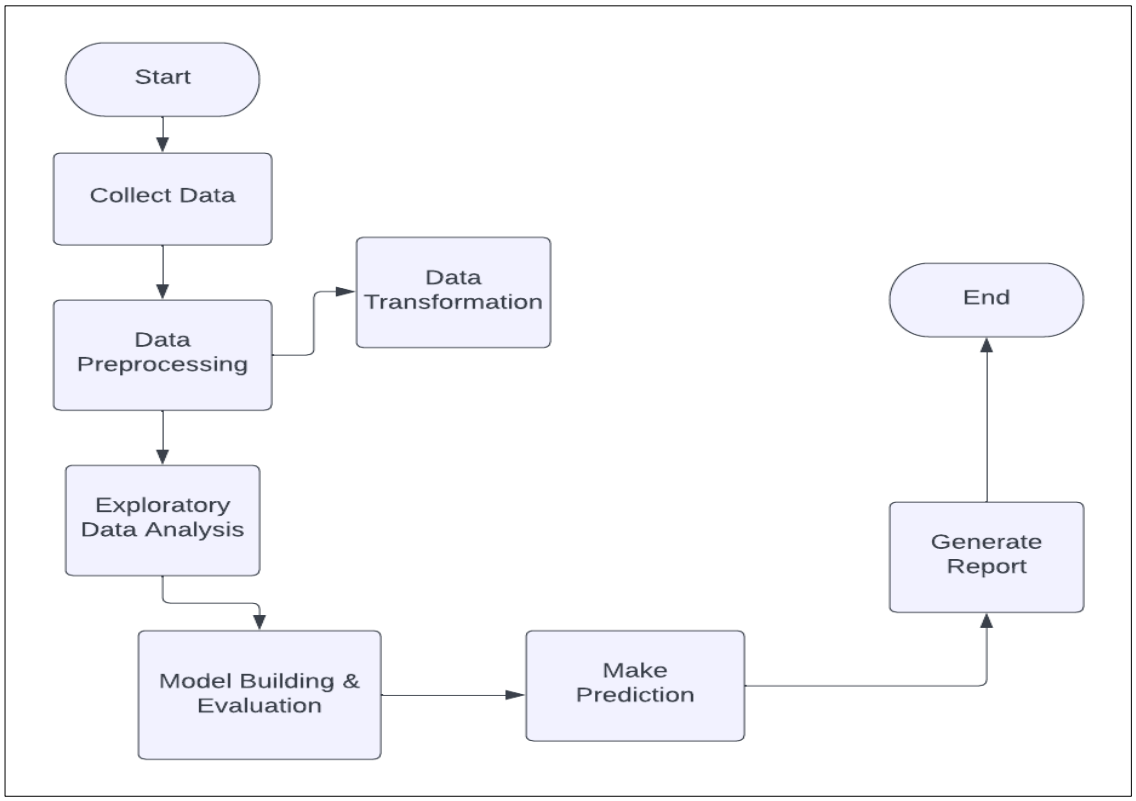
Within project management, planning is the method of creating the appropriate papers to ensure that the project is finished effectively. All responsibilities for defining, preparing, integrating, and coordinating further plans have been documented. The project plan outlines how the project will be implemented, managed, and concluded.

Project planning necessitates a thorough examination and organization of the following activities:

- The study was started on 15th March 2022.
- Finalizing the problem statement and research question was on 17th March 2022.
- Introduction chapter was started writing on 21st March 2022 and sources for the second chapter was started on 25th March 2022.
- After that, theoretical foundation was written till 07th May 2022.
- The third chapter research methodology was written and its sources collected till 18th June 2022.
- Data collection started on 10th June and analysis and finding results done on 18th July 2022.
- Conclusion chapter was completed on 03rd August 2022 and final checking finished till 11th August 2022.
- On the 16th August 2022, thesis is submitted.

3.4.2 Project flow

Figure 24: Representation of project flow



Source: Author's own representation

Flowcharts include graphs for creating and documenting simple systems and procedures. They like other types of diagrams, help visualize what's really happening and thus assist in comprehending a process, plus possibly discovering fewer features like flaws and constraints. The project flow of this paper is as follows:

- After deciding problem statement, research questions, and objective of the paper, start collecting data for the study.
- Better data collection was mandatory for data pre-processing and after that transformation of data was conducted.
- Exploratory data analysis was beginning to provide better visualisation of data and give insights from the data.
- Apply Bayesian optimization to make predictions.
- After that make a report by suggesting shortest possible way and end this project.

3.4.3 Data Collection and Data Description

The data set on which this code is written was obtained from Kaggle. This data set contains aggregate stopped vehicle information and intersection wait times at intersections in the major US cities like Atlanta, Boston, Chicago and Philadelphia in 2019. There are two shape of data sets displayed which are train data set and test data set. Train data set has 15 more columns than test data set and test data set has twice the number of observations. The main principal components of data set are:

Intersection: An intersection is an at-grade junction where two or more roads or streets meet or cross.

EntryStreetName: It is the street name where the traffic or vehicle enters the intersection.

ExitStreentName: It is the street name where the traffic or vehicle exiting the intersection.

EntryHeading and ExitHeading: Those are the directions the vehicle was traveling when entering/exiting the intersection.

TotalTimeStopped_p20,_p40,_p50,_p60,_p80 : It represents the total stopped time or waiting time by percentiles.

DistanceToFirstStop: DistanceToFirstStop is how far before the intersection the vehicle stopped for the first time.

TimeFromFirstStop: TimeFromFirstStop is how long it took from that point to cross the intersection.

3.4.4 Data Exploration

After importing all the necessary packages, the data injection is finished with this coding:

Figure 25: Importing Data

```
'''Read in export and import data from CSV file'''
df_train = pd.read_csv('../input/bigquery-geotab-intersection-congestion/train.csv')
df_test = pd.read_csv('../input/bigquery-geotab-intersection-congestion/test.csv')
```

Source: Author's own representation

After the injection of database, we need to get the glimpse of data, the following code was used:

Figure 26: Description of data

```
'''Dimension of train and test data'''
bold('**Shape of our train and test data**')
print('Dimension of train:',df_train.shape)
print('Dimension of test:',df_test.shape)
```

Source: Author's own representation

After the code was run the output says that the shape of train dataset has 856387 rows and 28 columns and shape of test dataset has 1921357 rows and 13 columns. Next, to get the gist of the data the following code was used and later the df. describe () function was used to get the sense of the data.

Figure 27: Description of data

```
'''Train and Test data at a glance.'''
bold('**Preview of Train Data:**')
display(df_train.head())
bold('**Preview of Test Data:**')
display(df_test.head())
```

Source: Author's own representation

Figure 28: Glimpse of data

	RowId	IntersectionId	Latitude	Longitude	EntryStreetName	ExitStreetName	EntryHeading	ExitHeading	Hour	Weekend	...	TimeFromFirstStop_p40	TimeFromFirstStop_p50
0	1921357	0	33.791659	-84.430032	Marietta Boulevard Northwest	Marietta Boulevard Northwest	NW	NW	0	0	...	0.0	0.0
1	1921358	0	33.791659	-84.430032	Marietta Boulevard Northwest	Marietta Boulevard Northwest	SE	SE	0	0	...	0.0	0.0
2	1921359	0	33.791659	-84.430032	Marietta Boulevard Northwest	Marietta Boulevard Northwest	NW	NW	1	0	...	0.0	0.0
3	1921360	0	33.791659	-84.430032	Marietta Boulevard Northwest	Marietta Boulevard Northwest	SE	SE	1	0	...	0.0	0.0
4	1921361	0	33.791659	-84.430032	Marietta Boulevard Northwest	Marietta Boulevard Northwest	NW	NW	2	0	...	0.0	0.0

Source: Author's own representation

Figure 29: Glimpse of data

	RowId	IntersectionId	Latitude	Longitude	EntryStreetName	ExitStreetName	EntryHeading	ExitHeading	Hour	Weekend	Month	Path	City
0	0	1	33.75094	-84.393032	Peachtree Street Southwest	Peachtree Street Southwest	NE	NE	0	0	6	Peachtree Street Southwest_NE_Peachtree Street...	Atlanta
1	1	1	33.75094	-84.393032	Peachtree Street Southwest	Mitchell Street Southwest	SW	SE	0	0	6	Peachtree Street Southwest_SW_Mitchell Street ...	Atlanta
2	2	1	33.75094	-84.393032	Peachtree Street Southwest	Peachtree Street Southwest	SW	SW	0	0	6	Peachtree Street Southwest_SW_Peachtree Street...	Atlanta
3	3	1	33.75094	-84.393032	Peachtree Street Southwest	Peachtree Street Southwest	NE	NE	1	0	6	Peachtree Street Southwest_NE_Peachtree Street...	Atlanta
4	4	1	33.75094	-84.393032	Peachtree Street Southwest	Peachtree Street Southwest	SW	SW	1	0	6	Peachtree Street Southwest_SW_Peachtree Street...	Atlanta

Source: Author's own representation

The above diagrams represent the head of the data set. The whole data set is being given in the appendix of this paper.

3.4.5 Algorithm Description

Bayesian optimization: Bayesian Optimization is a systematic method which is based on the Bayes Theorem that directs an efficient and successful search of a global optimization issue.[129] It functions by creating a surrogate function, a probabilistic model of the objective function, which is then effectively searched with an acquisition function before candidate samples are chosen for evaluation on the actual objective function.

In machine learning, Bayesian optimization is frequently used to fine-tune the hyperparameters of a particular well-performing model on a validation dataset.

3.4.6 Tools used in study

Python for coding:

Python is a powerful programming language with a high level of abstraction. Its design ethos prioritizes the readability of the code and makes extensive usage indentation.[130] Python was designed in 1980 by Guido van Rossum at the Netherlands' National Research Institute for Mathematics and Computer Science as a successor to the ABC language that could handle exceptions and interfaces.[131] On either a small or big scale, the force allows the user to build clear programs. Python's biggest notable feature is that it supports a variety of programming paradigms, including object-oriented, imperative, functional, and iterative programming. Python features a vast and extensive standard library, as well as a dynamic type system and automatic memory management.

Features of Python:

Python has a variety of beneficial features that distinguish it from many other programming languages.

---

[129] Cp. *Martin Pelikan, David E Goldberg, and Erick Cantú-Paz.*, The Bayesian Optimization Algorithm, no page number.
[130] Cp. *John Kramer and K R Srinath.*, The Fastest Growing Programming Language, 2017, no page number.
[131] Cp. *Arun Kumar and Supriya P. Panda.*, Python Pitches in IT-World, 2019, pp. 248–51.

- Simple to Understand and Apply: Python is a simple programming useful language when compared to other programming languages. Its syntax is simple and similar to that of the English language.[132]
- Language That Conveys: Python is capable of performing complex tasks with just a few lines of code. A basic example is the welcome home program, which can create by typing and then do print ("Welcome Home").[133]
- Accessibility of languages: Python has the authority to empower or connect MATLAB, R, FORTRAN, and C, C++ libraries and functions from Python using MLabWrap, RPy, F2pyCtypes, Cython, and SWIG, respectively.
- Object-Oriented Programming (OOP): Python mostly employs classes to implement the OOP paradigm. Class inheritance is common, however, some of them, such as diamond inheritance, can cause problems with class initialization.[134]
- Assistance for GUI Programming: For the development of a desktop program, a graphical user interface is used. The libraries used to construct the web service include PyQT5, Tkinter, and Kivy.
- Organization of Data: Python's adaptive service allows different data structures to exist.

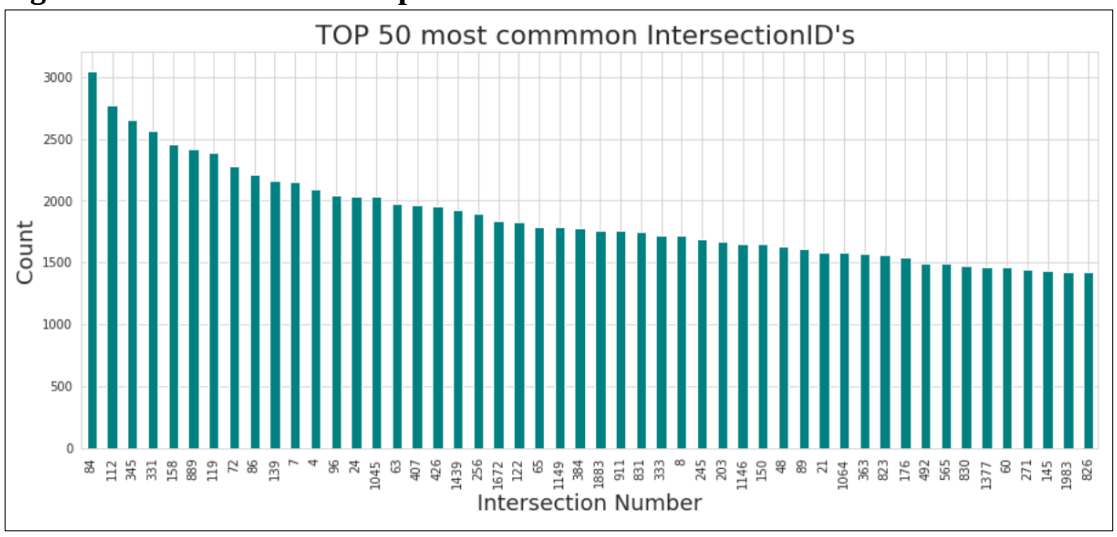
[132] Cp. *D Lakshminarayanan and S Prabhakaran.*, Python Programming Language, 2020, no page number.
[133]See https://books.google.de/books?hl=en&lr=&id=mh0bU6NXrBgC&oi=fnd&pg=PR1&dq=what+are+features+of+python+programming+language&ots=XBjGyu5Zjb&sig=lKt94lX2d0NcuXodYMU_j5hX2S0&redir_esc=y#v=onepage&q=what are features of python programming language&f=false, accessed March 22, 2022; *Wesley Chun*, Core Python Programming, no page number.
[134] Cp. *Yun Peng, Yu Zhang, and Mingzhe Hu.*, An Empirical Study, 2021, pp. 24–35.

4. Research Results

4.1 Data Modelling, Evaluation, and Deployment

The first visualisation is about top 50 most common intersection id from the data.

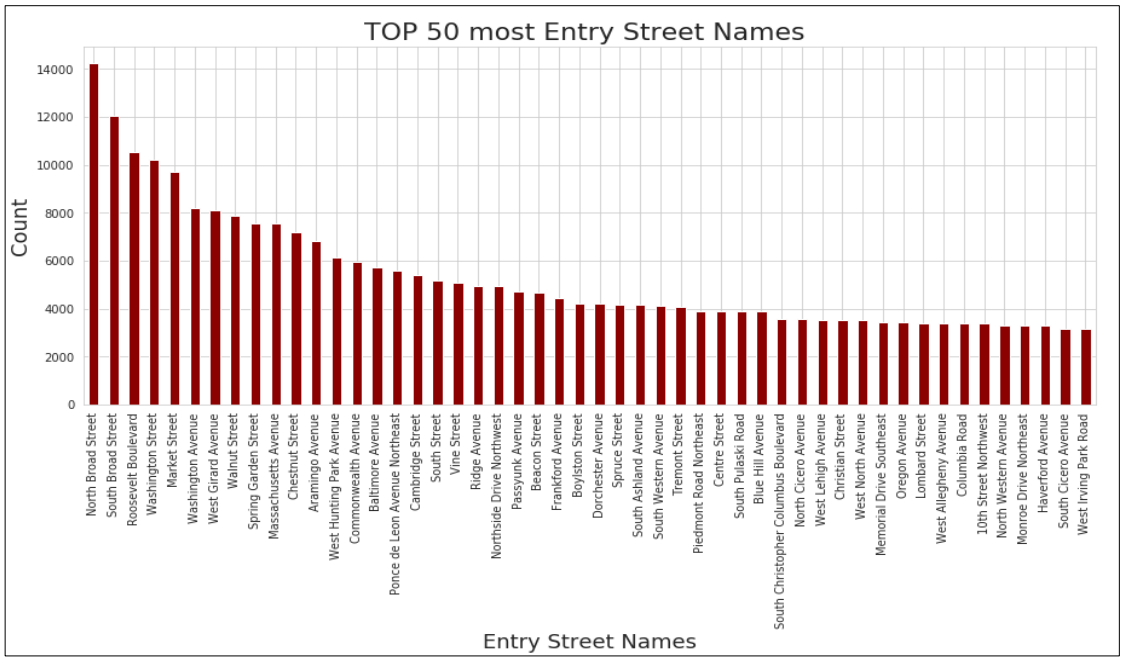
Figure 30: Visualisation of top 50 most common intersection id's



Source: Author's own representation

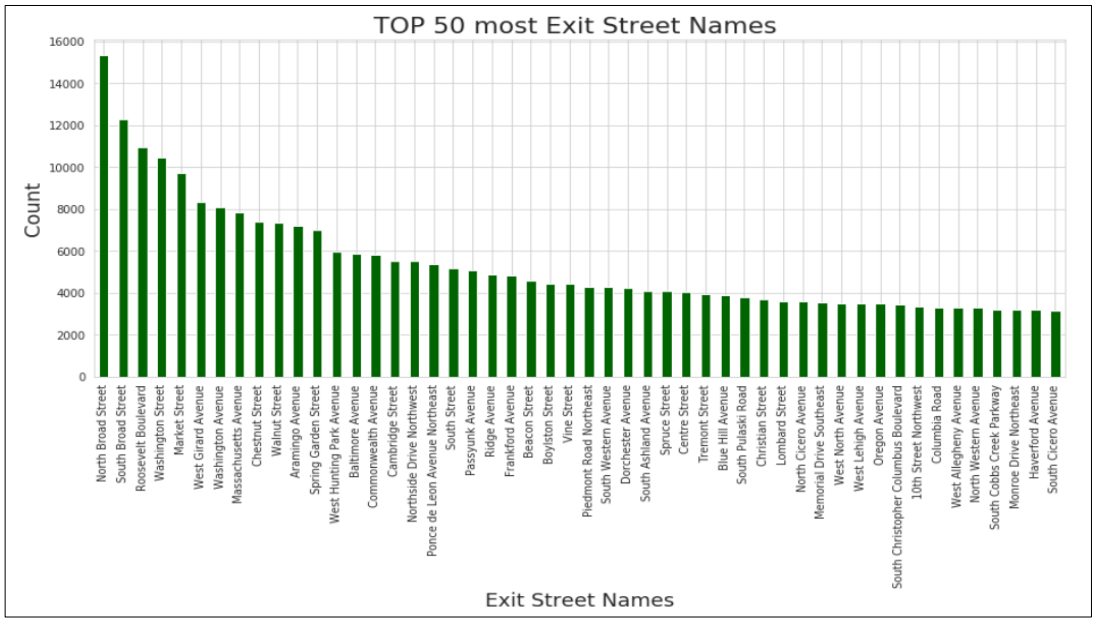
After that visualisation of top 50 most entry and exit street names is carried out to find out which street has more traffic.

Figure 31: Visualisation of top 50 entry street names



Source: Author's own representation

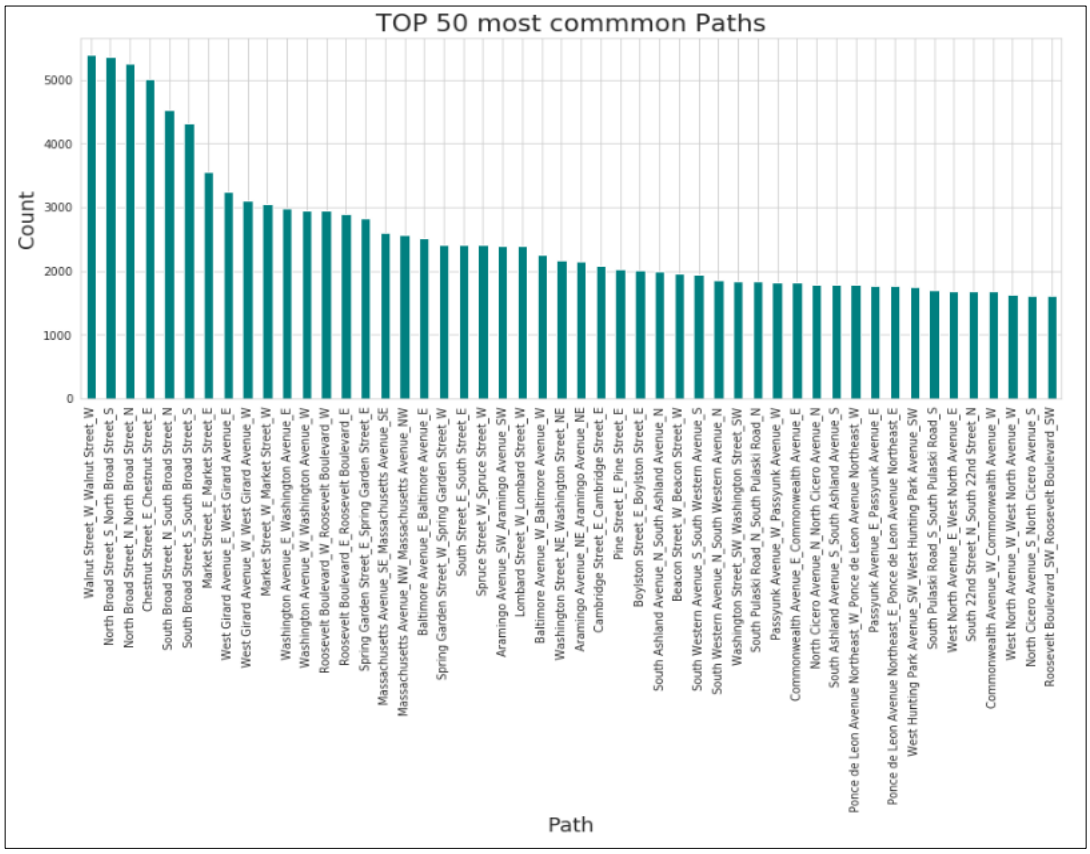
Figure 32: Visualisation of top 50 exit street names



Source: Author's own representation

Then visualisation of top 50 common paths as follows:

Figure 33: Visualisation of top 50 most common paths

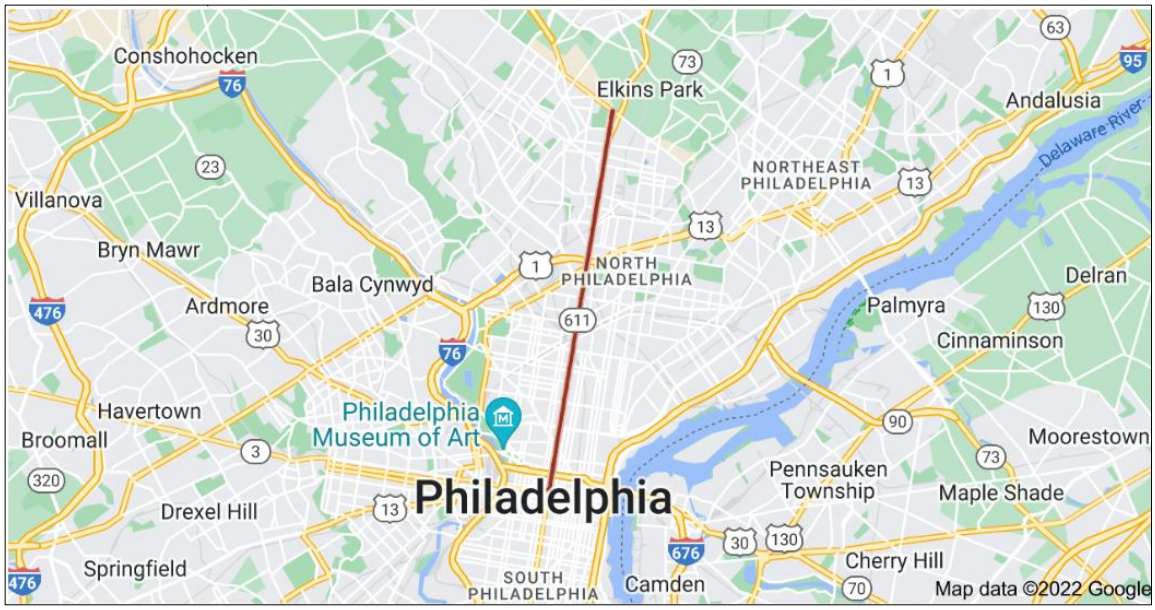


Source: Author's own representation

From the above visualisation, North Broad Street is the most Entry and Exit Street Names.

As seen in the below map the most common path is North Broad Street_N.

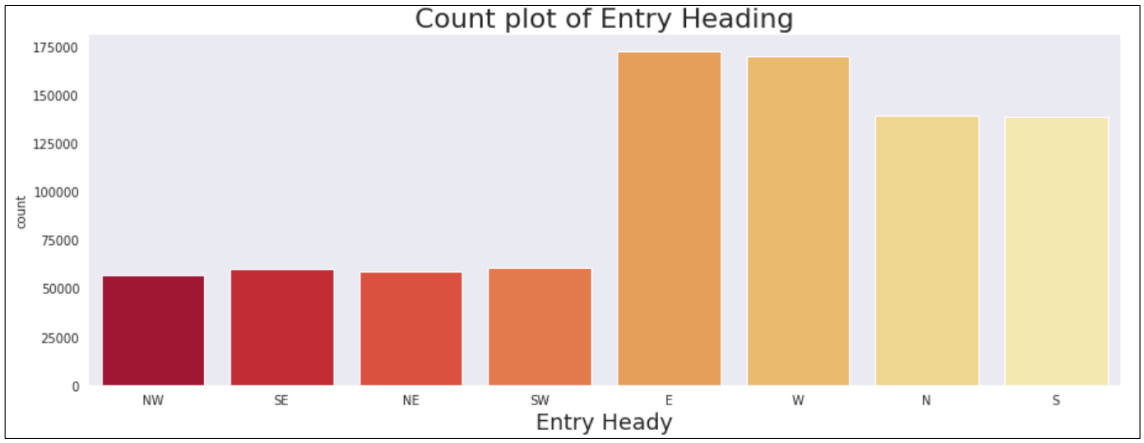
Figure 34: Visualisation of most common path



Source: Author's own representation

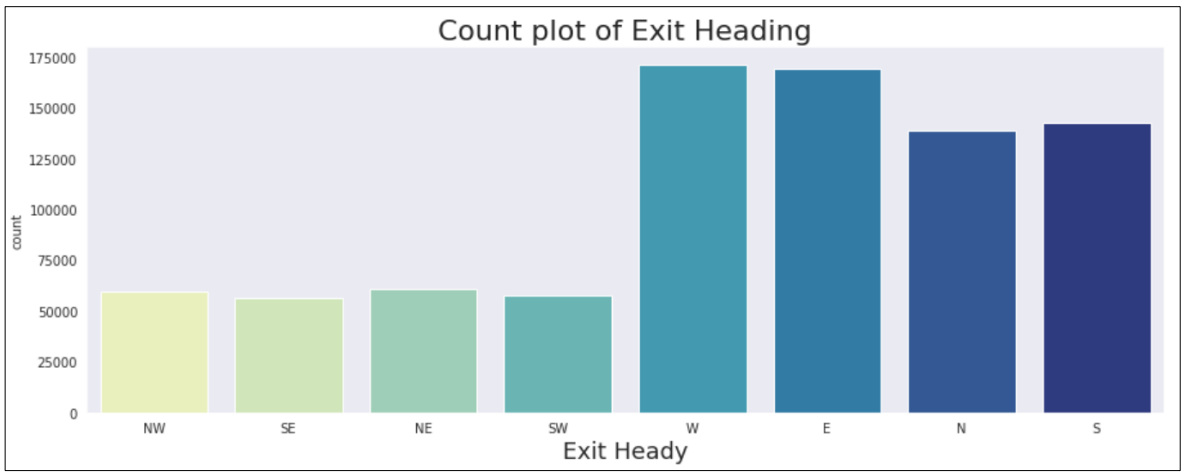
The visualisation of entry/exit heading is carried out.

Figure 35: Bar plot of entry heading



Source: Author's own representation

Figure 36: Bar plot of exit heading

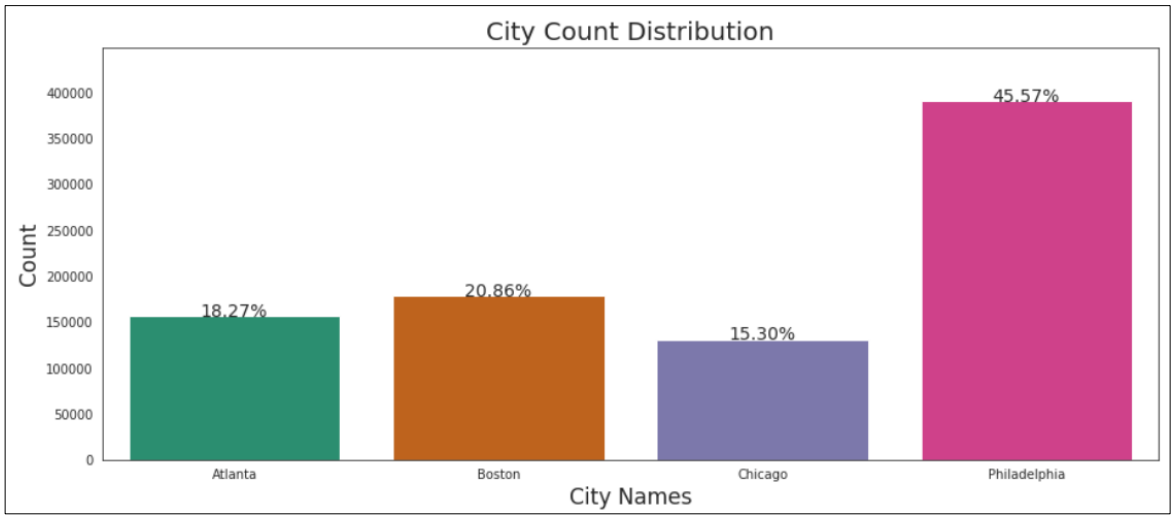


Source: Author's own representation

From the bar chart, we can observe that in general Entry and Exit Heading is exactly the same and most of the vehicle are heading toward East and West.

The analysis of city and time feature is carried together to get clear picture about the data and its distribution.

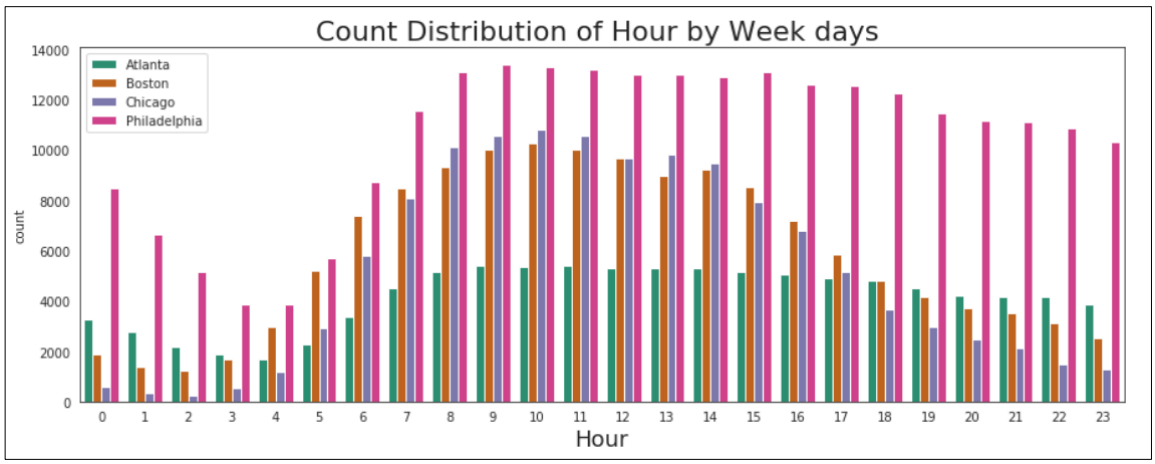
Figure 37: Visualisation of city count distribution



Source: Author's own representation

In the city count distribution, Philadelphia have 45.29% of the total entries followed by Boston (21.23%), Atlanta (17.89%), Chicago (15.59%).

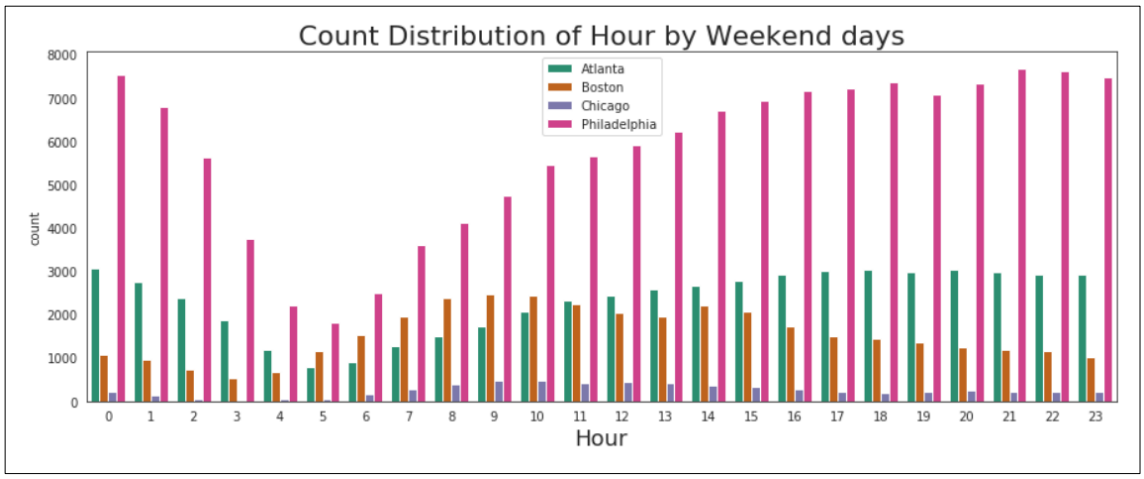
Figure 38: Visualisation of count distribution of hour by weekdays



Source: Author's own representation

In the Distribution of Hour by Week days, between 08:00 to 17:00, is the rush hour in all cities, but for Philadelphia, it is 08:00 to 24:00.

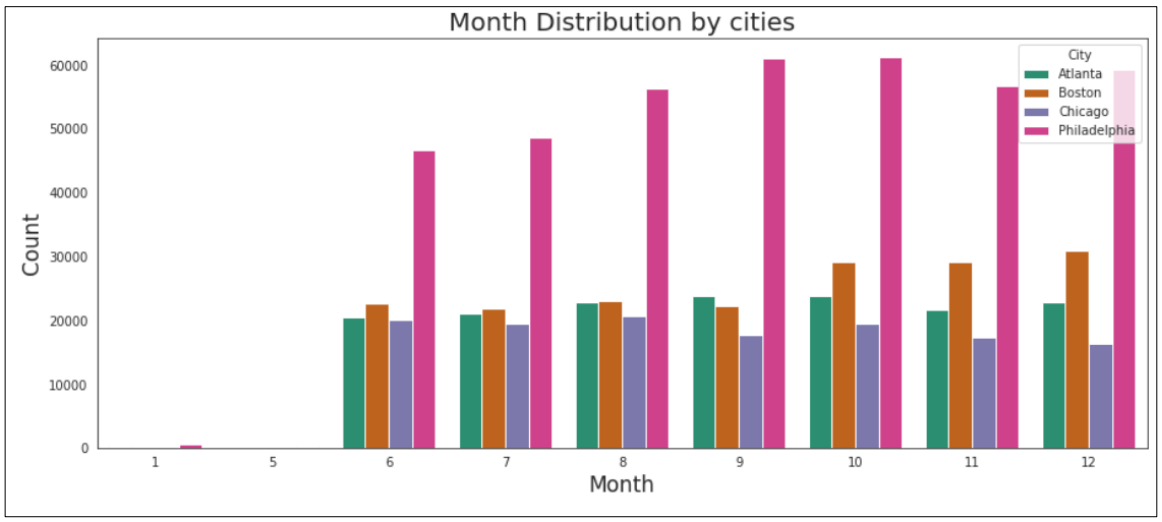
Figure 39: Visualisation of count distribution of hour by weekend days



Source: Author's own representation

In the Distribution of Hour by Weekend days, all the cities are by far the most common in all hours during the weekend, only on 05:00 to 7:00 hour.

Figure 40: Bar plot of month distribution by cities

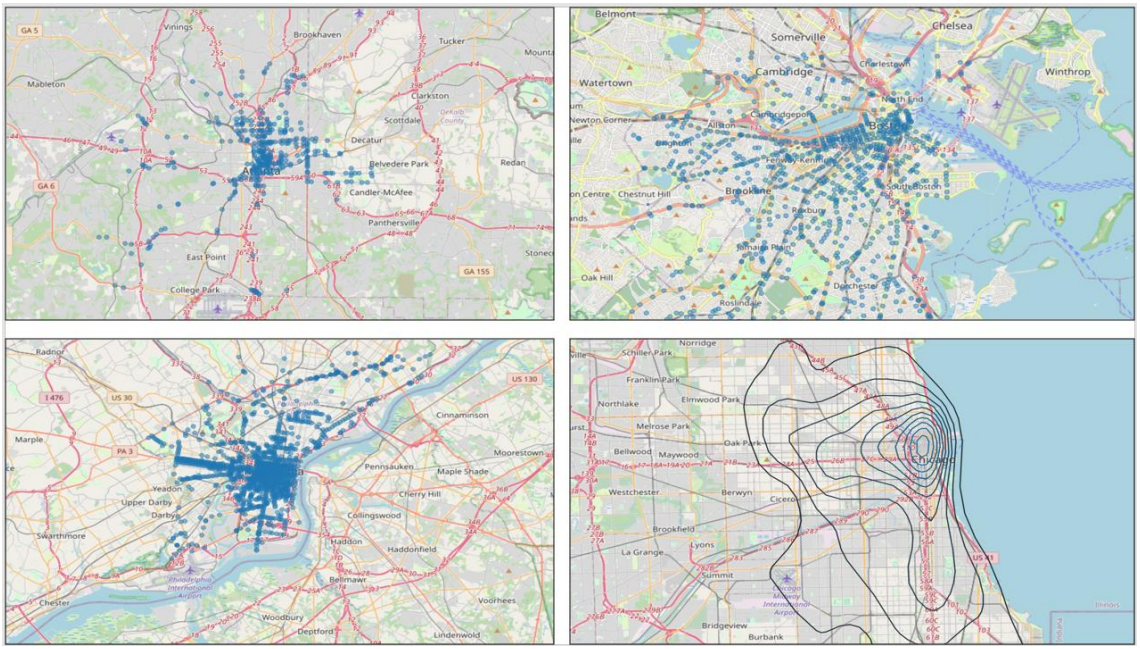


Source: Author's own representation

In the Distribution of month by cities, it's seemed that we have only six months of data only.

The below plots are visualisation map plot of 4 cities in US.

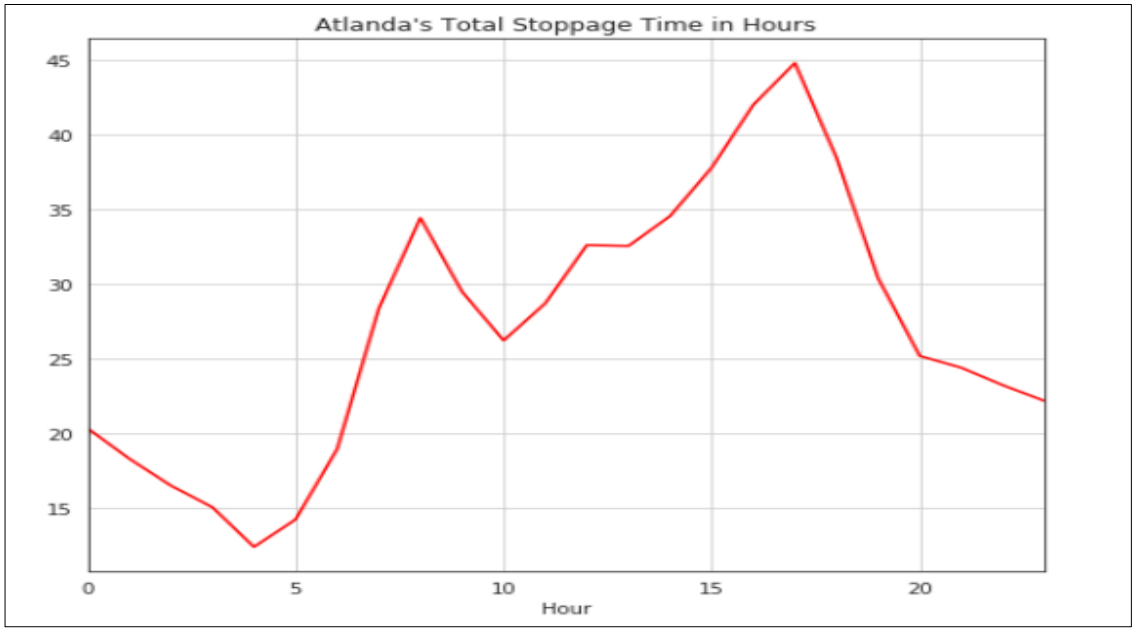
Figure 41: Map visualisation of 4 cities



Source: Author's own representation

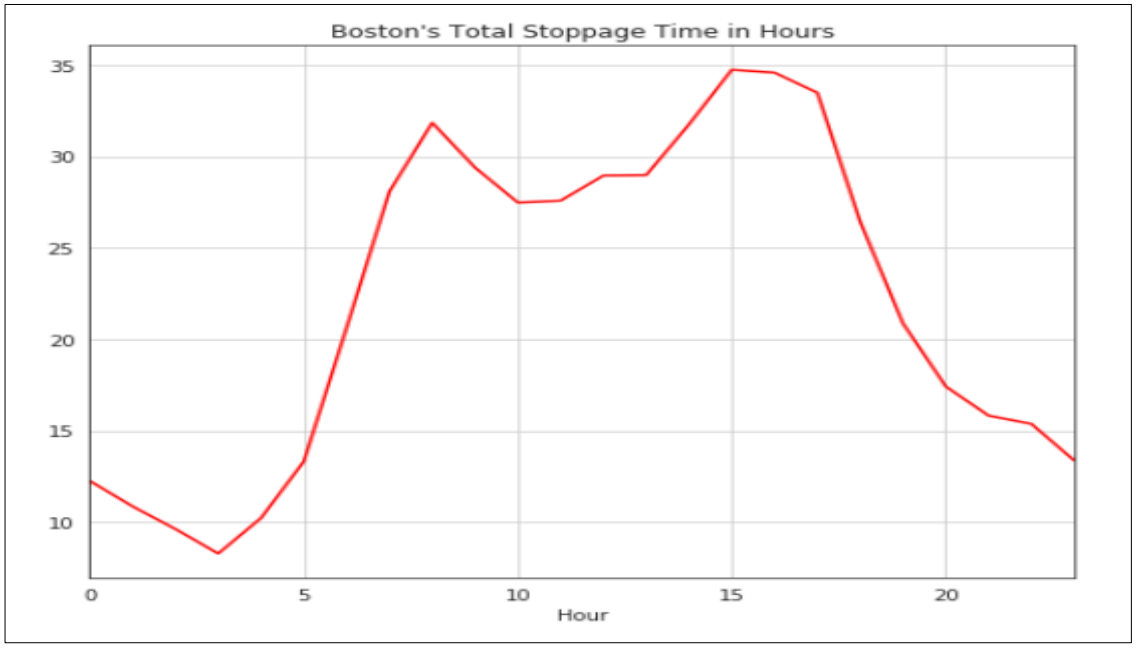
After above analysis, in the next section, we are going analysis the total time stopped on the intersections in different cities.

Figure 42: Representation of stop time of Atlanta



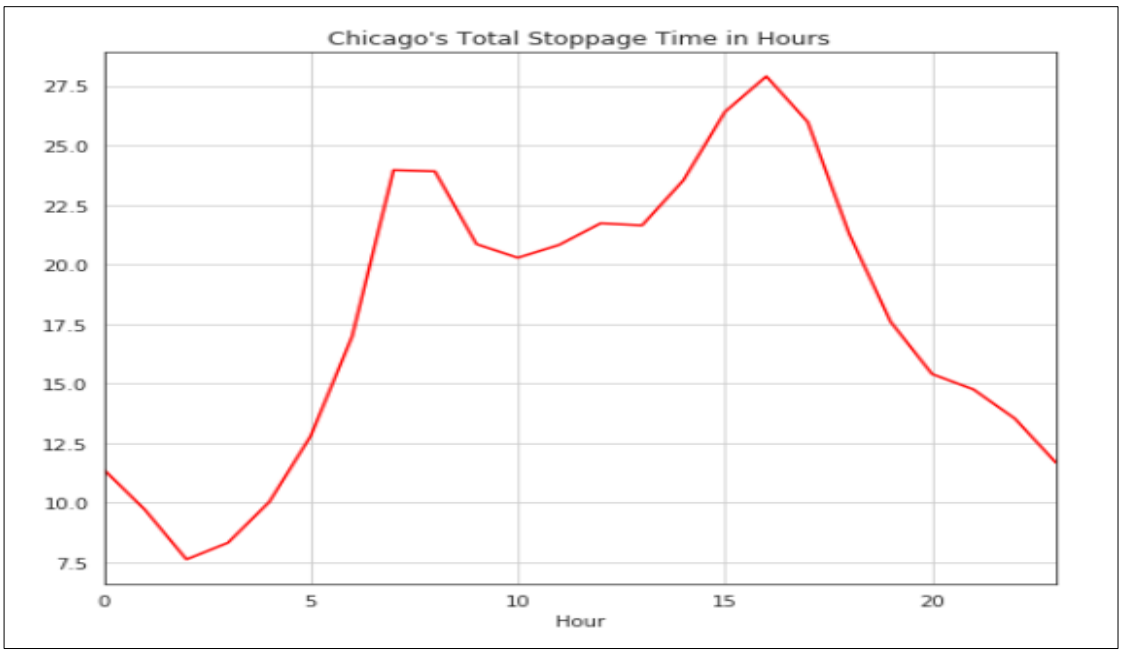
Source: Author's own representation

Figure 43: Representation of stop time of Boston



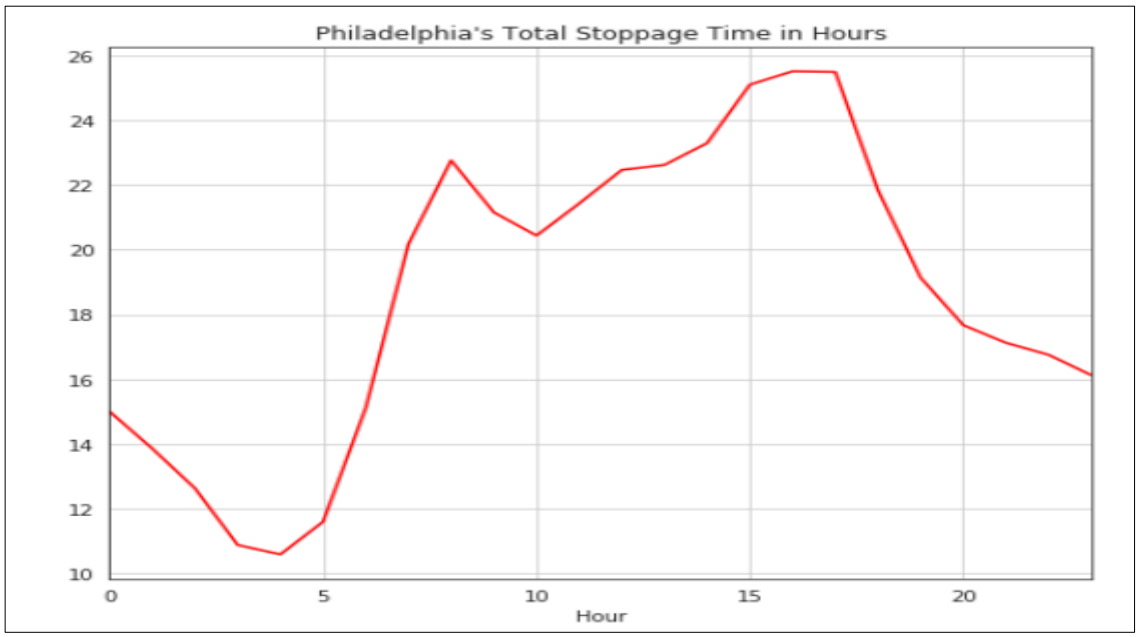
Source: Author's own representation

Figure 44: Representation of stop time of Chicago



Source: Author's own representation

Figure 45: Representation of stop time of Philadelphia



Source: Author's own representation

The above time plot shows that during the evening from 15:00 to 17:00 stoppage time has significantly increased.

Encoding street name is performed for prediction. Encoding turn direction is then carried out. The cardinal directions can be expressed using the equation:

Formula 1: Equation for the cardinal directions

$$\theta/\pi \quad (1)$$

Source: Based on Spiros Skiadopoulos, M., Encoding turn, 2001, p. 300

Where, θ is the angle between the direction we want to encode and the north compass direction, measured clockwise.

Figure 46: Representation of encoding street names

	ExitHeading	EntryHeading	diffHeading
0	1.75	1.75	0.0
1	0.75	0.75	0.0
381	0.75	0.25	-0.5
654	0.75	1.25	0.5
655	0.25	0.75	0.5

Source: Author's own representation

4.2 Results and Findings

After implementing Bayesian optimization, define objective function and then define surrogate model of the objective function and called it optimizer.

Figure 47: Representation of surrogate model

iter	target	baggin...	featur...	lambda_l1	lambda_l2	max_depth	min_ch...	min_sp...	num_le...
1	-6.751	0.967	0.1419	2.234	1.803	11.95	12.58	0.03998	153.2
2	-6.241	0.902	0.3841	2.882	2.283	12.7	3.156	0.02257	224.5
3	-6.059	0.9887	0.4527	1.939	1.069	15.01	3.644	0.04641	213.6
4	-6.298	0.8517	0.3659	2.588	0.7442	14.19	4.024	0.01812	113.5
5	-6.385	0.8406	0.2821	2.384	4.954	16.86	8.171	0.05206	107.7
6	-6.744	0.927	0.1294	2.412	4.46	13.93	17.39	0.001582	177.0
7	-6.729	0.8091	0.1651	1.243	0.06253	17.3	19.81	0.02974	218.0
8	-6.214	0.8	0.5	3.0	0.0	8.0	1.0	0.1	193.7
9	-6.905	1.0	0.1	0.0	5.0	19.0	1.0	0.1	199.1
10	-6.127	0.9427	0.4082	2.133	2.401	15.81	4.571	0.06461	213.1
11	-6.226	0.8	0.5	0.0	0.0	8.0	20.0	0.1	122.5
12	-6.569	0.8254	0.2304	2.892	0.5225	8.056	18.17	0.06197	100.2
13	-6.236	0.8	0.5	0.0	5.0	8.0	1.0	0.1	132.1
14	-6.216	1.0	0.5	3.0	0.0	8.0	20.0	0.1	200.3
15	-6.229	1.0	0.5	0.0	5.0	8.0	1.0	0.001	212.5
16	-6.092	0.8	0.5	3.0	5.0	19.0	20.0	0.1	129.5
17	-6.905	1.0	0.1	0.0	0.0	19.0	1.0	0.001	131.9

Source: Author's own representation

The below values are prediction of entry heading before the business junction point where users can take easy path to avoid congestion.

Figure 48: Representation of prediction of path

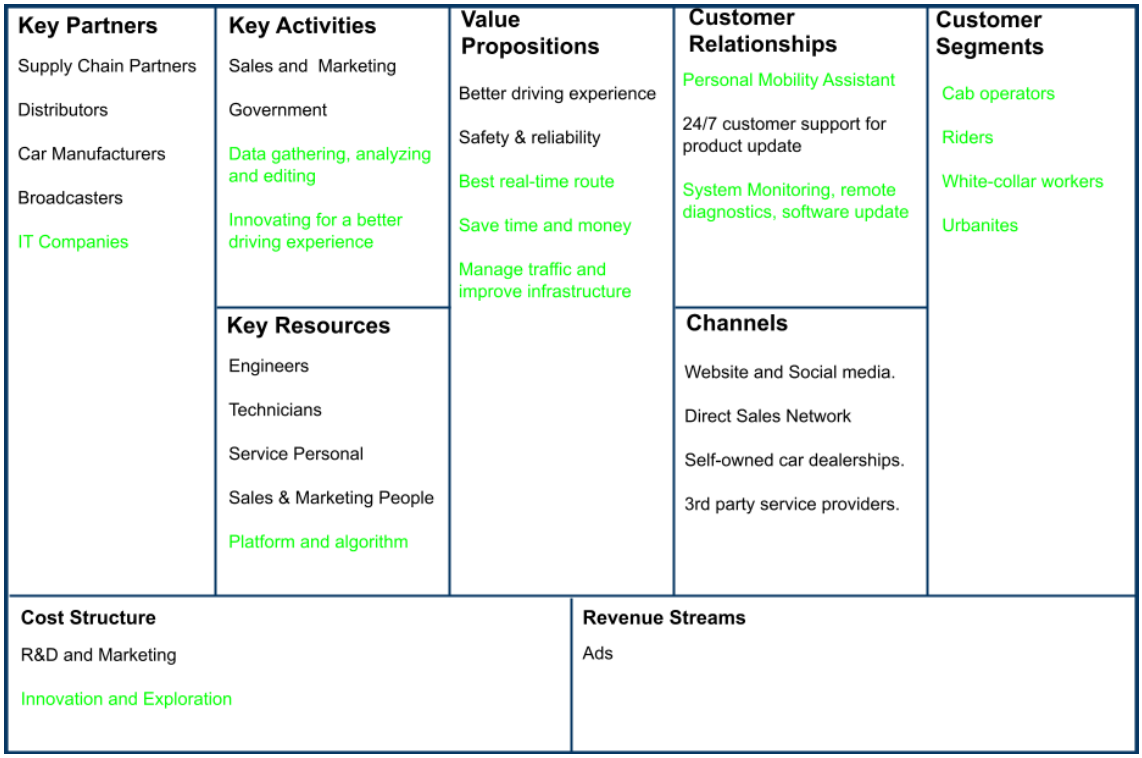
	TargetId	Target
0	0_0	0.872841
1	0_1	2.557215
2	0_2	11.630482
3	0_3	4.872446
4	0_4	7.837346

Source: Author's own representation

5. Projection of Results

The Business Model Canvas for the Automobile Industry which is helpful for this project is as follows:

Figure 49: Representation of BMC of Automobile Industry

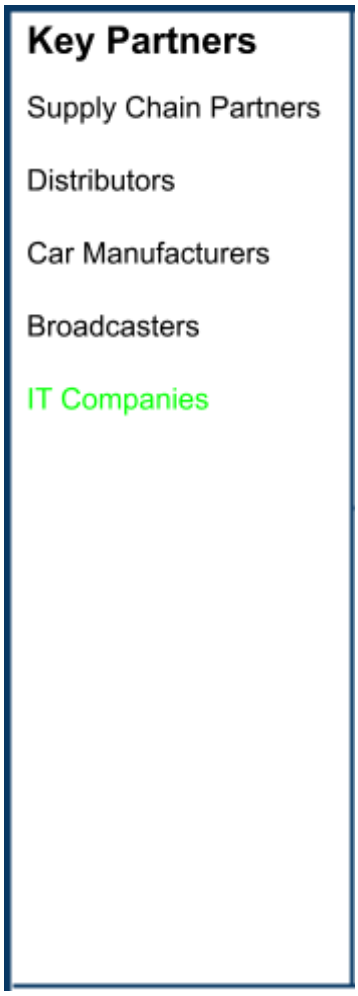


Source: Author's own representation

This BMC comprises Automobile Industry in which the algorithm is community-driven for users to provide best possible and fastest route information. Basically, there are nine components in this BMC.

In this BMC, IT companies are key partners for this algorithm by providing support in the Automobile Industry.

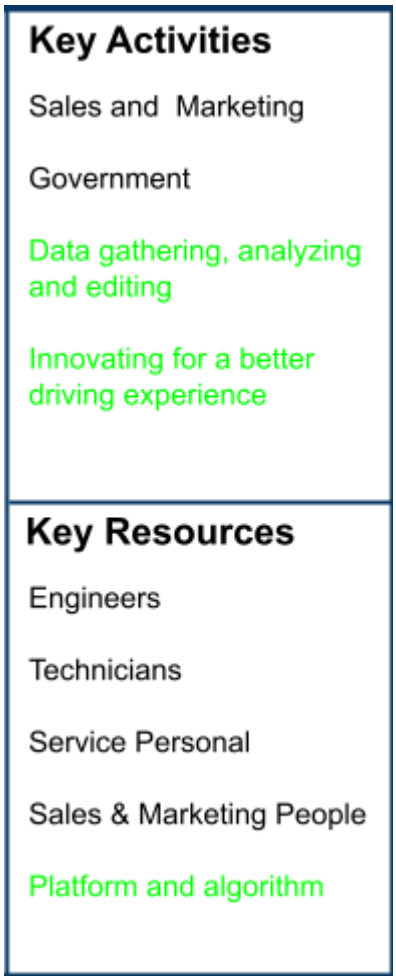
Figure 50: Representation of key partners in BMC



Source: Author's own representation

Data gathering, data analysing and editing and innovation for better driving experience are the key activity in this paper. Additionally, platform and algorithm are key resources for this BMC.

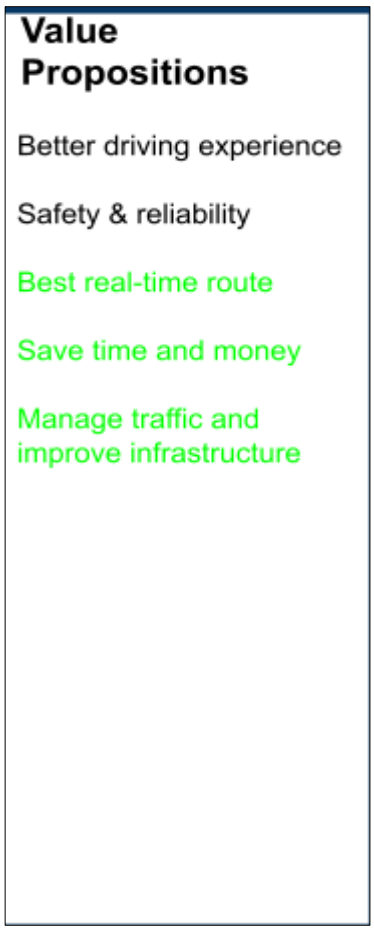
Figure 51: Representation of key activities & key resources of BMC



Source: Author's own representation

Value proposition for this BMC is to provide best real-time route to the users. Furthermore, saving time and money of the users are adding value to this BMC. By doing this it is very helpful in managing traffic and improve infrastructure of the BMC.

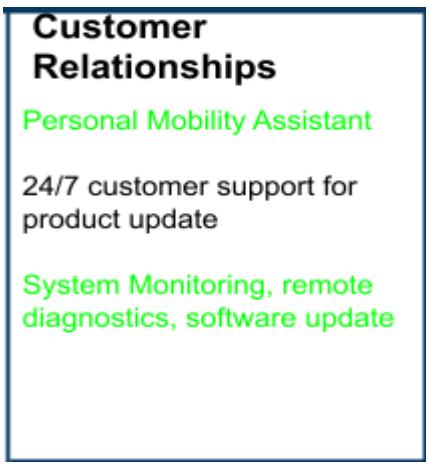
Figure 52: Representation of value proposition of BMC



Source: Author's own representation

Customer relationship is built by giving personal mobility assistance to users and proving them system monitoring, remote diagnostics and software update.

Figure 53: Representation of customer relationships in BMC



Source: Author's own representation

Figure 54: Representation of customer segments in BMC

Source: Author's own representation

In this BMC, the target audience is cab drivers, riders, white-collar workers and urbanites who are busy and have to reach their destination as soon as possible.

Figure 55: Representation of cost structure in BMC

Source: Author's own representation

In this BMC, for innovation and exploration cost is necessity for the better performance.

6. Conclusion

The objective of this paper was to development of an AI/ML algorithm which will reduce factors like incremental delay, vehicle operation cost, fuel operating, and cost emission by suggesting the shortest possible route for drivers. Driving is a means of travelling that is preferred by numerous people. Individuals want to drive efficiently and reach their destination with as little time stuck in traffic jam as possible. This paper gather data about traffic congestion of 4 major US cities and visualise several factors. There are 4 major US cities such as Atlanta, Boston, Chicago, and Philadelphia. After visualisation, we find out that Philadelphia has more traffic compared to other three cities. It also analyses rush hours in all three cities from 8 in the morning to 5 in the evening but for Philadelphia from 8 in the morning to 12 in the midnight. Moreover, due to heavy traffic the stoppage time from 15 to 17 hour during evening has significantly rose in all cities.

The entry and exit heading in this paper have been encoded into numbers so that the model can use encoding for the prediction. Thus, model predicted the entry heading before the business junction point where driver can easily go to avoid the intersection where traffic is jammed.

Limitations:

There are some drawbacks of this study in Automobile Industry as follows:

- The data is obtained from Kaggle so it is limited data which is only six months of data.
- Weather and temperature affect traffic.
- Urban locations: likely if a street is numbered but Rural locations are not, may affect the traffic.
- The intersection is surrounded by businesses: may affect the traffic during weeks and weekend day.

Future Scope:

There are some applications of this paper in the automobile industry which could be helpful for the industry

- Apply ensemble learning for better predictive performance.

- Use this algorithm on next stage and create an application which is more user friendly and give better performance.
- Collect more detailed data and include several other factors for better results.

Appendix

This section contains the code used for predictive models for congestion, based on an aggregate measure of stopping distance and waiting times, at intersections in 4 major US cities: Atlanta, Boston, Chicago & Philadelphia.

```
'''Read in export and import data from CSV file'''
df_train = pd.read_csv('../input/bigquery-geotab-intersection-congestion/train.csv')
df_test = pd.read_csv('../input/bigquery-geotab-intersection-congestion/test.csv')
```

```
'''Train and Test data at a glance.'''
bold('**Preview of Train Data:**')
display(df_train.head())
bold('**Preview of Test Data:**')
display(df_test.head())
```

```
'''Dimension of train and test data'''
bold('**Shape of our train and test data**')
print('Dimension of train:',df_train.shape)
print('Dimension of test:',df_test.shape)
```

```
'''Funtion for Variable Description'''
def description(df):
    summary = pd.DataFrame(df.dtypes,columns=['dtypes'])
    summary = summary.reset_index()
    summary['Name'] = summary['index']
    summary = summary[['Name','dtypes']]
    summary['Missing'] = df.isnull().sum().values
    summary['Uniques'] = df.nunique().values
    summary['First Value'] = df.iloc[0].values
    summary['Second Value'] = df.iloc[1].values
    summary['Third Value'] = df.iloc[2].values
    return summary
```

```
bold('**Variable Description of  train Data:**')
display(description(df_train))
bold('**Variable Description of  test Data:**')
display(description(df_test))
```

```
'''Visulization of IntersectionID'''
plt.figure(figsize=(15,6))
df_train.IntersectionId.value_counts()[:50].plot(kind='bar', color = 'teal')
plt.xlabel("Intersection Number", fontsize=18)
plt.ylabel("Count", fontsize=18)
plt.title("TOP 50 most commmon IntersectionID's ", fontsize=22)
plt.show()
```

```
'''Visulization of Entry/Exit StreetNames'''
plt.figure(figsize=(15,6))
df_train['EntryStreetName'].value_counts()[:50].plot(kind='bar', color = 'darkred')
plt.xlabel("Entry Street Names", fontsize=18)
plt.ylabel("Count", fontsize=18)
plt.title("TOP 50 most Entry Street Names ", fontsize=22)
plt.show()

plt.figure(figsize=(15,6))
df_train['ExitStreetName'].value_counts()[:50].plot(kind='bar', color = 'darkgreen')
plt.xlabel("Exit Street Names", fontsize=18)
plt.ylabel("Count", fontsize=18)
plt.title("TOP 50 most Exit Street Names ", fontsize=22)
plt.show()
```

```
'''Visulization of Path'''
plt.figure(figsize=(15,6))
df_train.Path.value_counts()[:50].plot(kind='bar', color = 'teal')
plt.xlabel("Path", fontsize=18)
plt.ylabel("Count", fontsize=18)
plt.title("TOP 50 most commmon Paths", fontsize=22)
plt.show()
```

```
sns.set_style("dark")
fig, ax = plt.subplots(2,1, figsize=[15, 12])

sns.countplot(data = df_train, x = 'EntryHeading', ax = ax[0], palette = 'YlOrRd_r')
ax[0].set_title('Count plot of Entry Heading', fontsize = 22)
ax[0].set_xlabel('Entry Heady', fontsize = 18)

sns.countplot(data = df_train, x = 'ExitHeading', ax = ax[1], palette = 'YlGnBu')
ax[1].set_title('Count plot of Exit Heading', fontsize = 22)
ax[1].set_xlabel('Exit Heady', fontsize = 18)

plt.subplots_adjust(hspace = 0.3)
plt.show()
```

```
total = len(df_train)
plt.figure(figsize=(15,6))
sns.set_style("white")

ax = sns.countplot(x = "City", data = df_train, palette = 'Dark2')
ax.set_title("City Count Distribution", fontsize=20)
ax.set_ylabel("Count",fontsize= 17)
ax.set_xlabel("City Names", fontsize=17)
sizes=[]
for p in ax.patches:
    height = p.get_height()
    sizes.append(height)
    ax.text(p.get_x()+p.get_width()/2.,
            height + 3,
            '{:1.2f}%'.format(height/total*100),
            ha="center", fontsize=14)
ax.set_ylim(0, max(sizes) * 1.15)

plt.show()
```

```
fig, ax = plt.subplots(2,1, figsize=[15, 12])

sns.countplot(data = df_train[df_train['Weekend']==0], x = 'Hour', hue = 'City', ax
= ax[0], palette = 'Dark2')
ax[0].legend()
ax[0].set_title('Count Distribution of Hour by Week days ', fontsize = 22)
ax[0].set_xlabel('Hour', fontsize = 18)

sns.countplot(data = df_train[df_train['Weekend']==1], x = 'Hour', hue = 'City', ax
= ax[1], palette = 'Dark2')
ax[1].legend()
ax[1].set_title('Count Distribution of Hour by Weekend days', fontsize = 22)
ax[1].set_xlabel('Hour', fontsize = 18)

plt.subplots_adjust(hspace = 0.3)
plt.show()
```

```
plt.figure(figsize=(15,6))
ax = sns.countplot(x = "Month", data = df_train, hue = 'City',palette = 'Dark2')
ax.set_title("Month Distribution by cities", fontsize=20)
ax.set_ylabel("Count",fontsize= 17)
ax.set_xlabel("Month", fontsize=17)
plt.show()
```

```
'''Visualition of Map Plot of Atlanta city'''
import mplleaflet
plt.figure(figsize=(10,10))
map1 = df_train[df_train['City']=='Atlanta'].groupby(['Latitude', 'Longitude'])['Row
Id'].count().reset_index()
plt.scatter(map1['Longitude'], map1['Latitude'], alpha=0.5)

mplleaflet.display()
```

```
'''Visualition of Map Plot of Boston city'''
plt.figure(figsize=(10,10))
map2 = df_train[df_train['City']=='Boston'].groupby(['Latitude', 'Longitude'])['RowI
d'].count().reset_index()
plt.scatter(map2['Longitude'], map2['Latitude'], alpha=0.5)

mplleaflet.display()
```

```
'''Visualition of Map Plot of Philadelphia city'''
plt.figure(figsize=(10,10))
map3 = df_train[df_train['City']=='Philadelphia'].groupby(['Latitude', 'Longitude'])
['RowId'].count().reset_index()
plt.scatter(map3['Longitude'], map3['Latitude'], alpha=0.5)

mplleaflet.display()
```

```
'''Visualition of Map Plot of Chicago city'''
plt.figure(figsize=[10, 10])
map4 = df_train[df_train['City']=='Chicago'].groupby(['Latitude', 'Longitude'])['Row
Id'].count().reset_index()
sns.kdeplot(map4['Longitude'], map4['Latitude'])

mplleaflet.display()
```

```
fig, ax = plt.subplots(nrows=2, ncols=2)
sns.set_style("whitegrid")

df_train[df_train['City']=='Atlanta'].groupby('Hour')['TotalTimeStopped_p80'].mean().plot(
    ax=ax[0,0],title="Atlanda's Total Stoppage Time in Hours", color='r', figsize=(18,15))

df_train[df_train['City']=='Boston'].groupby('Hour')['TotalTimeStopped_p80'].mean().plot(
    ax=ax[0,1],title="Boston's Total Stoppage Time in Hours", color='r', figsize=(18,15))


df_train[df_train['City']=='Chicago'].groupby('Hour')['TotalTimeStopped_p80'].mean().plot(
    ax=ax[1,0],title="Chicago's Total Stoppage Time in Hours", color='r', figsize=(18,15))


df_train[df_train['City']=='Philadelphia'].groupby('Hour')['TotalTimeStopped_p80'].mean().plot(
    ax=ax[1,1],title="Philadelphia's Total Stoppage Time in Hours", color='r', figsize=(18,15))

plt.show()
```

```
'''Let's use the following road types: Street, Avenue, Road, Boulevard, Broad and Drive'''
road_encoding = {
    'Road': 1,
    'Street': 2,
    'Avenue': 2,
    'Drive': 3,
    'Broad': 3,
    'Boulevard': 4
}
```

```
def encode(x):
    if pd.isna(x):
        return 0
    for road in road_encoding.keys():
        if road in x:
            return road_encoding[road]

    return 0
```

```
df_train['EntryTypeStreet'] = df_train['EntryStreetName'].apply(encode)
df_train['ExitTypeStreet'] = df_train['ExitStreetName'].apply(encode)
df_test['EntryTypeStreet'] = df_test['EntryStreetName'].apply(encode)
df_test['ExitTypeStreent'] = df_test['ExitStreetName'].apply(encode)
print(df_train['EntryTypeStreet'].sample(10).values)
```

```
df_train["same_street_exact"] = (df_train["EntryStreetName"] ==  df_train["ExitStreetName"]).astype(int)
df_test["same_street_exact"] = (df_test["EntryStreetName"] ==  df_test["ExitStreetName"]).astype(int)
```

```
directions = {
    'N': 0,
    'NE': 1/4,
    'E': 1/2,
    'SE': 3/4,
    'S': 1,
    'SW': 5/4,
    'W': 3/2,
    'NW': 7/4
}
```

```
df_train['EntryHeading'] = df_train['EntryHeading'].map(directions)
df_train['ExitHeading'] = df_train['ExitHeading'].map(directions)

df_test['EntryHeading'] = df_test['EntryHeading'].map(directions)
df_test['ExitHeading'] = df_test['ExitHeading'].map(directions)

df_train['diffHeading'] = df_train['EntryHeading']- df_train['ExitHeading']
df_test['diffHeading'] = df_test['EntryHeading']- df_test['ExitHeading']

display(df_train[['ExitHeading','EntryHeading','diffHeading']].drop_duplicates().head(5))
```

```
import lightgbm as lgb
from bayes_opt import BayesianOptimization

dtrain = lgb.Dataset(data=X_train, label=y1_train)

'''Define objective function'''
def hyp_lgbm(num_leaves, feature_fraction, bagging_fraction, max_depth, min_split_gain, min_child_weight, lambda_
l1, lambda_l2):

        params = {'application':'regression','num_iterations': 400,
                  'learning_rate':0.01,
                  'metric':'rmse'} # Default parameters
        params["num_leaves"] = int(round(num_leaves))
        params['feature_fraction'] = max(min(feature_fraction, 1), 0)
        params['bagging_fraction'] = max(min(bagging_fraction, 1), 0)
        params['max_depth'] = int(round(max_depth))
        params['min_split_gain'] = min_split_gain
        params['min_child_weight'] = min_child_weight
        params['lambda_l1'] = lambda_l1
        params['lambda_l2'] = lambda_l2

        cv_results = lgb.cv(params, dtrain, nfold=5, seed=44, categorical_feature=cat_feat, stratified=False,
                            verbose_eval =None)
#         print(cv_results)
        return -np.min(cv_results['rmse-mean'])
```

```
pds = {'num_leaves': (100, 230),
           'feature_fraction': (0.1, 0.5),
           'bagging_fraction': (0.8, 1),
           'lambda_l1': (0,3),
           'lambda_l2': (0,5),
           'max_depth': (8, 19),
           'min_split_gain': (0.001, 0.1),
           'min_child_weight': (1, 20)
         }
```

```
optimizer = BayesianOptimization(hyp_lgbm,pds,random_state=44)

# Optimize
optimizer.maximize(init_points=5, n_iter=12)
```

```
'''Best parameters after optimization'''
optimizer.max
```

```
submission = pd.read_csv('../input/bigquery-geotab-intersection-congestion/sample_submission.csv')
#submission.head()
```

```
dt = pd.DataFrame(all_preds).stack()
dt = pd.DataFrame(dt)
submission['Target'] = dt[0].values
```

```
submission.head()
```

Bibliography

3 Philosophy-Based Research: The Entry of Quantitative and Qualitative Models into Educational Research | River Publishers Books | IEEE Xplore. accessed February 28, 2022. https://ieeexplore.ieee.org/document/9645516

A Framework for Strict Structured Literature Reviews within Management Research and Management Education: Establishing a Model for Evidence-Based Management and Increased Scholarly Interaction, 2018

Aghion, Philippe, Céline Antonin, and Simon Bunel. "Artificial Intelligence, Growth and Employment: The Role of Policy." *Economie et Statistique* 2019, no. 510–512 (2019): 149–64. https://doi.org/10.24187/ecostat.2019.510t.1994

Alloghani, Mohamed, Dhiya Al-Jumeily, Jamila Mustafina, Abir Hussain, and Ahmed J. Aljaaf. "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," 2020, 3–21. https://doi.org/10.1007/978-3-030-22475-2_1

Alur, Rajeev, and Radu Grosu. "Modular Refinement of Hierarchic Reactive Machines." *Conference Record of the Annual ACM Symposium on Principles of Programming Languages*, 2000, 390–402. https://doi.org/10.1145/325694.325746

Andresen, Scott L. "John McCarthy: Father of Al." *IEEE Intelligent Systems* 17, no. 5 (2002): 84–85. https://doi.org/10.1109/MIS.2002.1039837

Benko, Attila, and Cecília Sik Lányi. "History of Artificial Intelligence." *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/978-1-60566-026-4.Ch276*, January 1, 1AD, 1759–62. https://doi.org/10.4018/978-1-60566-026-4.CH276

Blocker, Christopher P., and Daniel J. Flint. "Customer Segments as Moving Targets: Integrating Customer Value Dynamism into Segment Instability Logic." *Industrial Marketing Management* 36, no. 6 (August 1, 2007): 810–22. https://doi.org/10.1016/J.INDMARMAN.2006.05.016

Booker, L. B., D. E. Goldberg, and J. H. Holland. "Classifier Systems and Genetic Algorithms." *Artificial Intelligence* 40, no. 1–3 (September 1, 1989): 235–82. https://doi.org/10.1016/0004-3702(89)90050-7

Brézillon, Patrick. “Context in Artificial Intelligence: I. A Survey of the Literature RUNNING TITLE: Context in Artificial Intelligence: A Survey,” n.d

Busetto, Loraine, Wolfgang Wick, and Christoph Gumbinger. “How to Use and Assess Qualitative Research Methods.” *Neurological Research and Practice* 2, no. 1 (December 2020). https://doi.org/10.1186/S42466-020-00059-Z

Cao, Hung, Monica Wachowicz, and Sangwhan Cha. “Developing an Edge Computing Platform for Real-Time Descriptive Analytics.” *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017* 2018-January (July 1, 2017): 4546–54. https://doi.org/10.1109/BIGDATA.2017.8258497

Carlos Niebles, Juan, Hongcheng Wang, Li Fei-Fei, JC Niebles, H Wang, and L Fei-Fei. “Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words.” *International Journal of Computer Vision 2008 79:3* 79, no. 3 (March 4, 2008): 299–318. https://doi.org/10.1007/S11263-007-0122-4

Casadesus-Masanell, Ramon, and Joan E Ricart. “HBR.ORG Spotlight on BuSineSS Model Innovation How to Design A Winning Business Model,” n.d

“Characterization of Classification Algorithms,” n.d

Chardonnens, Thibaud, Philippe Cudre-Mauroux, Martin Grund, and Benoit Perroud. “Big Data Analytics on High Velocity Streams: A Case Study.” *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 2013, 784–87. https://doi.org/10.1109/BIGDATA.2013.6691653

Chesbrough, Henry. “Business Model Innovation: Opportunities and Barriers.” *Long Range Planning* 43, no. 2–3 (April 1, 2010): 354–63. https://doi.org/10.1016/J.LRP.2009.07.010

Collins, Christopher, Denis Dennehy, Kieran Conboy, and Patrick Mikalef. “Artificial Intelligence in Information Systems Research: A Systematic Literature Review and Research Agenda.” *International Journal of Information Management* 60 (October 1, 2021): 102383. https://doi.org/10.1016/J.IJINFOMGT.2021.102383

Connelly, Lynne M. “Inclusion and Exclusion Criteria.” *MEDSURG Nursing* 29, no. 2 (2020): 125. https://doi.org/10.4324/9780203340653_chapter_6

Cools, Mario, Elke Moons, Brecht Janssens, and Geert Wets. “Shifting towards Environment-Friendly Modes: Profiling Travelers Using Q-Methodology.” *Transportation* 36, no. 4 (April 24, 2009): 437–53. https://doi.org/10.1007/S11116-009-9206-Z/TABLES/3

Covarrubias, Alex. “When Disruptors Converge: The Last Automobile Revolution.” *International Journal of Automotive Technology and Management* 18, no. 2 (2018): 81–104. https://doi.org/10.1504/IJATM.2018.092184

Crisci, C., B. Ghattas, and G. Perera. “A Review of Supervised Machine Learning Algorithms and Their Applications to Ecological Data.” *Ecological Modelling* 240 (August 10, 2012): 113–22. https://doi.org/10.1016/J.ECOLMODEL.2012.03.001

Dayan, Peter, and Yael Niv. “Reinforcement Learning: The Good, The Bad and The Ugly.” *Current Opinion in Neurobiology* 18, no. 2 (April 1, 2008): 185–96. https://doi.org/10.1016/J.CONB.2008.08.003

Dibb, Sally, and Lyndon Simkin. “Judging the Quality of Customer Segments: Segmentation Effectiveness.” *Http://Dx.Doi.Org/10.1080/09652540903537048* 18, no. 2 (April 2010): 113–31. https://doi.org/10.1080/09652540903537048

Duan, Yanqing, John S. Edwards, and Yogesh K. Dwivedi. “Artificial Intelligence for Decision Making in the Era of Big Data – Evolution, Challenges and Research Agenda.” *International Journal of Information Management* 48 (October 1, 2019): 63–71. https://doi.org/10.1016/J.IJINFOMGT.2019.01.021

Elgendy, Nada, and Ahmed Elragal. “Big Data Analytics: A Literature Review Paper.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8557 LNAI (2014): 214–27. https://doi.org/10.1007/978-3-319-08976-8_16/COVER/

Engelen, Jesper E. van, and Holger H. Hoos. “A Survey on Semi-Supervised Learning.” *Machine Learning* 109, no. 2 (February 1, 2020): 373–440. https://doi.org/10.1007/S10994-019-05855-6/FIGURES/5

Ezenwoye, Onyeka. “What Language? - The Choice of an Introductory Programming Language.” *Proceedings - Frontiers in Education Conference, FIE* 2018-October (March 4, 2019). https://doi.org/10.1109/FIE.2018.8658592

Faisal, Asif, Tan Yigitcanlar, Graham Currie, Source Journal, Land Use, and Tan Yigitcanlar. “Understanding Autonomous Vehicles Linked References Are Available on JSTOR for This Article : Understanding Autonomous Vehicles : A Systematic Literature Re- View on Capability , Impact , Planning and Policy.” *Journal of Transport and Land Use* 12, no. 1 (2019): 45–72

“Frequently Used Classification Algorithms in Machine Learning with Comparative Analysis of Various Parameters | International Journal of Advanced Science and Technology.” Accessed June 24, 2022. http://sersc.org/journals/index.php/IJAST/article/view/27585

Frith, Chris, and Uta Frith. “Theory of Mind.” *Current Biology : CB* 15, no. 17 (2005): 644–45. https://doi.org/10.1016/j.cub.2005.08.041

Gandomi, Amir, and Murtaza Haider. “Beyond the Hype: Big Data Concepts, Methods, and Analytics.” *International Journal of Information Management* 35, no. 2 (April 1, 2015): 137–44. https://doi.org/10.1016/J.IJINFOMGT.2014.10.007

Gobble, Mary Anne M. “Big Data: The Next Big Thing in Innovation.” *Http://Dx.Doi.Org/10.5437/08956308X5601005* 56, no. 1 (January 2015): 64–66. https://doi.org/10.5437/08956308X5601005

Goertzel, Ben. “Artificial General Intelligence: Concept, State of the Art, and Future Prospects.” *Journal of Artificial General Intelligence* 5, no. 1 (2014): 2013–15. https://doi.org/10.2478/jagi-2014-0001

Guo, Yanming, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. “Deep Learning for Visual Understanding: A Review.” *Neurocomputing* 187 (April 26, 2016): 27–48. https://doi.org/10.1016/J.NEUCOM.2015.09.116

Gupta, Gunjan K., Alexander Strehl, and Joydeep Ghosh. “Distance Based Clustering of Association Rules.” *Intelligent Engineering Systems Through Artificial Neural Networks* 9, no. section IV (1999): 759–64

Hsu, Wei Chun, Lin Lin, and Chen Yu Li. “Forecasting Automobile Sales: The Peña-Box Approach.” *Transportation Planning and Technology* 37, no. 6 (2014): 568–80. https://doi.org/10.1080/03081060.2014.921408

Ishwarappa, and J. Anuradha. “A Brief Introduction on Big Data 5Vs Characteristics and

Hadoop Technology." *Procedia Computer Science* 48, no. C (2015): 319–24. https://doi.org/10.1016/J.PROCS.2015.04.188

Iyamu, Tiko. "South African Journal of Information Management Affiliation," 2020. https://doi.org/10.4102/sajim.v22i1.1117

Jazwinski, Andrew H. "Limited Memory Optimal Filtering." *IEEE Transactions on Automatic Control* 13, no. 5 (1968): 558–63. https://doi.org/10.1109/TAC.1968.1098981

Jet, Akinsola, and Hinmikaiye J O. "Supervised Machine Learning Algorithms: Classification and Comparison." *International Journal of Computer Trends and Technology* 48 (2017). https://doi.org/10.14445/22312803/IJCTT-V48P126

Jiang, Tammy, Jaimie L. Gradus, and Anthony J. Rosellini. "Supervised Machine Learning: A Brief Primer." *Behavior Therapy* 51, no. 5 (September 1, 2020): 675–87. https://doi.org/10.1016/J.BETH.2020.05.002

Kang, Yue, Zhao Cai, Chee Wee Tan, Qian Huang, and Hefu Liu. "Natural Language Processing (NLP) in Management Research: A Literature Review." *Https://Doi.Org/10.1080/23270012.2020.1756939* 7, no. 2 (April 2, 2020): 139–72. https://doi.org/10.1080/23270012.2020.1756939

Kaur, Harkiran, and Aanchal Phutela. "Commentary upon Descriptive Data Analytics." *Proceedings of the 2nd International Conference on Inventive Systems and Control, ICISC 2018*, June 27, 2018, 678–83. https://doi.org/10.1109/ICISC.2018.8398884

Khan, M. Ali Ud Din, Muhammad Fahim Uddin, and Navarun Gupta. "Seven V's of Big Data Understanding Big Data to Extract Value." *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education - "Engineering Education: Industry Involvement and Interdisciplinary Trends", ASEE Zone 1 2014*, 2014. https://doi.org/10.1109/ASEEZONE1.2014.6820689

Khan, Nawsher, Mohammed Alsaqer, Habib Shah, Gran Badsha, Aftab Ahmad Abbasi, and Soulmaz Salehian. "The 10 Vs, Issues and Challenges of Big Data," 2018. https://doi.org/10.1145/3206157.3206166

Khanum Tahira Mahboob Assistant Professor Assistant Professor, Memoona, and Warda Imtiaz Humaraia Abdul Ghafoor Rabeea Sehar. "A Survey on Unsupervised

Machine Learning Algorithms for Automation, Classification and Maintenance." *International Journal of Computer Applications* 119, no. 13 (2015): 975–8887. www.campsearch.com

Kindström, Daniel, and Mikael Ottosson. "Local and Regional Energy Companies Offering Energy Services: Key Activities and Implications for the Business Model." *Applied Energy* 171 (June 1, 2016): 491–500. https://doi.org/10.1016/J.APENERGY.2016.03.092

Kramer, John, and K R Srinath. "Python-The Fastest Growing Programming Language." *International Research Journal of Engineering and Technology*, 2017. www.irjet.net

Kumar, Arun, and Supriya P. Panda. "A Survey: How Python Pitches in IT-World." *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, February 1, 2019, 248–51. https://doi.org/10.1109/COMITCON.2019.8862251

Kumar, Vaibhav, and M. L. "Predictive Analytics: A Review of Trends and Techniques." *International Journal of Computer Applications* 182, no. 1 (July 16, 2018): 31–37. https://doi.org/10.5120/IJCA2018917434

Kuusi, Osmo, and Sirkka Heinonen. "Scenarios From Artificial Narrow Intelligence to Artificial General Intelligence—Reviewing the Results of the International Work/Technology 2050 Study:" *Https://Doi.Org/10.1177/19467567221101637* 2022, no. 1 (May 18, 2022): 194675672211016. https://doi.org/10.1177/19467567221101637

Lakshminarayanan, D, and S Prabhakaran. "A STUDY ON PYTHON PROGRAMMING LANGUAGE." *Dogo Rangsang Research Journal UGC Care Group I Journal* 10, no. 12 (2020)

Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning." *Nature 2015 521:7553* 521, no. 7553 (May 27, 2015): 436–44. https://doi.org/10.1038/nature14539

Lepenioti, Katerina, Alexandros Bousdekis, Dimitris Apostolou, and Gregoris Mentzas. "Prescriptive Analytics: Literature Review and Research Challenges." *International*

Journal of Information Management 50 (February 1, 2020): 57–70. https://doi.org/10.1016/J.IJINFOMGT.2019.04.003

Li, Yu Feng, and De Ming Liang. "Safe Semi-Supervised Learning: A Brief Introduction." *Frontiers of Computer Science 2019 13:4* 13, no. 4 (June 18, 2019): 669–76. https://doi.org/10.1007/S11704-019-8452-2

Li, Yuxi. "DEEP REINFORCEMENT LEARNING: AN OVERVIEW," 2018. https://arxiv.org/abs/

Lu, Yutong, Depei Qian, Haohuan Fu, and Wenguang Chen. "Will Supercomputers Be Super-Data and Super-AI Machines?" *Communications of the ACM* 61, no. 11 (November 1, 2018): 82–87. https://doi.org/10.1145/3239556

Mangindaan, Michelle C Burger, and Evelyn Hendriana. "Creating Customer Trust through Customer Relationship Management in SMEs," no. March (2013)

Masuda, Taizo, Kenji Araki, Kenichi Okumura, Shinich Urabe, Yuki Kudo, Kazutaka Kimura, Takashi Nakado, Akinori Sato, and Masafumi Yamaguchi. "Next Environment-Friendly Cars: Application of Solar Power as Automobile Energy Source." *Conference Record of the IEEE Photovoltaic Specialists Conference* 2016-November (November 18, 2016): 580–84. https://doi.org/10.1109/PVSC.2016.7749663

Mcquaid, Ronald W. "The Theory of Partnership: Why Have Partnerships?" Accessed February 28, 2022. https://www.researchgate.net/publication/291300642

Mittal, Shweta, and Om Prakash Sangwan. "Big Data Analytics Using Machine Learning Techniques." *Proceedings of the 9th International Conference On Cloud Computing, Data Science and Engineering, Confluence 2019*, 2019, 203–7. https://doi.org/10.1109/CONFLUENCE.2019.8776614

Muggleton, Stephen. "Alan Turing and the Development of Artificial Intelligence." *AI Communications* 27, no. 1 (January 1, 2014): 3–10. https://doi.org/10.3233/AIC-130579

Myers, J Paul. "The Japanese Fifth Generation Computing Project: Curricular Applications." *J. Comput. Sci. Coll.* 36, no. 5 (2021): 77–86

Naeem, Muhammad, Tauseef Jamal, Jorge Diaz-Martinez, Shariq Aziz Butt, Nicolo Montesano, Muhammad Imran Tariq, Emiro De-la-Hoz-Franco, and Ethel De-La-Hoz-Valdiris. “Trends and Future Perspective Challenges in Big Data.” *Smart Innovation, Systems and Technologies* 253 (2022): 309–25. https://doi.org/10.1007/978-981-16-5036-9_30/COVER/

Nair, Anoop C., S. Ramalingam, and Ashvini Ravi. “Knowledge Creation within the Automobile Industry:” *Http://Dx.Doi.Org/10.5772/61090* 7 (January 1, 2015). https://doi.org/10.5772/61090

Najafabadi, Maryam M., Flavio Villanustre, Taghi M. Khoshgoftaar, Naeem Seliya, Randall Wald, and Edin Muharemagic. “Deep Learning Applications and Challenges in Big Data Analytics.” *Journal of Big Data 2015 2:1* 2, no. 1 (February 24, 2015): 1–21. https://doi.org/10.1186/S40537-014-0007-7

Narayanan, Uma, Varghese Paul, and Shelbi Joseph. “Different Analytical Techniques for Big Data Analysis: A Review.” *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017*, June 19, 2018, 372–82. https://doi.org/10.1109/ICECDS.2017.8390139

Nilsson, Nils J. “Artificial Intelligence: A Modern Approach.” *Artificial Intelligence* 82, no. 1–2 (April 1996): 369–80. https://doi.org/10.1016/0004-3702(96)00007-0

O’Leary, Daniel E. “Artificial Intelligence and Big Data.” *IEEE Intelligent Systems* 28, no. 2 (2013): 96–99. https://doi.org/10.1109/MIS.2013.39

Ooi, Melanie. “Future Trend in IM: The Smarter Car.” *IEEE Instrumentation and Measurement Magazine* 22, no. 2 (April 1, 2019): 33–34. https://doi.org/10.1109/MIM.2019.8674632

Orsato, R. J., and P. Wells. “The Automobile Industry & Sustainability.” *Journal of Cleaner Production* 15, no. 11–12 (January 1, 2007): 989–93. https://doi.org/10.1016/J.JCLEPRO.2006.05.035

U-Turn: The Rise and Demise of the Automobile Industry.” *Journal of Cleaner Production* 15, no. 11–12 (2007): 994–1006. https://doi.org/10.1016/j.jclepro.2006.05.019

Osterwalder, Alexander, Yves Pigneur, and Christopher L. Tucci. “Clarifying Business

Models: Origins, Present, and Future of the Concept." *Communications of the Association for Information Systems* 16, no. 1 (2005): 1. https://doi.org/10.17705/1CAIS.01601

Ovans, Andrea. "What Is a Business Model?," 2015

Parra, Eugenio, Christos Dimou, Juan Llorens, Valentín Moreno, and Anabel Fraga. "A Methodology for the Classification of Quality of Requirements Using Machine Learning Techniques." *Information and Software Technology* 67 (November 1, 2015): 180–95. https://doi.org/10.1016/J.INFSOF.2015.07.006

Payne, Adrian, Pennie Frow, and Andreas Eggert. "The Customer Value Proposition: Evolution, Development, and Application in Marketing," n.d. https://doi.org/10.1007/s11747-017-0523-z

Pelikan, Martin, David E Goldberg, and Erick Cantú-Paz. "BOA: The Bayesian Optimization Algorithm," n.d

Peng, Yun, Yu Zhang, and Mingzhe Hu. "An Empirical Study for Common Language Features Used in Python Projects." *Proceedings - 2021 IEEE International Conference on Software Analysis, Evolution and Reengineering, SANER 2021*, March 1, 2021, 24–35. https://doi.org/10.1109/SANER50967.2021.00012

Praveena, Anto, and B. Bharathi. "A Survey Paper on Big Data Analytics." *2017 International Conference on Information Communication and Embedded Systems, ICICES 2017*, October 17, 2017. https://doi.org/10.1109/ICICES.2017.8070723

Pulabaigari, Viswanath, Eswara Reddy, Padmanabha Reddy, P Viswanath, and B Eswara Reddy. "Semi-Supervised Learning: A Brief Review Offline Handwritten Signature Identification and Verification View Project Synthesizing Handwritten Characters View Project Semi-Supervised Learning: A Brief Review." *International Journal of Engineering &Technology* 7, no. 1 (2018): 81–85. https://doi.org/10.14419/ijet.v7i1.8.9977

Qiu, Junfei, Qihui Wu, Guoru Ding, Yuhua Xu, and Shuo Feng. "A Survey of Machine Learning for Big Data Processing." *Eurasip Journal on Advances in Signal Processing* 2016, no. 1 (December 1, 2016): 1–16. https://doi.org/10.1186/S13634-016-0355-X/FIGURES/5

Reddy Nadikattu, Rahul. “THE EMERGING ROLE OF ARTIFICIAL INTELLIGENCE IN MODERN SOCIETY.” *International Journal of Creative Research Thoughts* 4 (2016): 2320–2882. www.ijcrt.orgwww.ijcrt.org

Reim, Wiebke, Sambit Lenka, Johan Frishammar, and Vinit Parida. “Implementing Sustainable Product–Service Systems Utilizing Business Model Activities.” *Procedia CIRP* 64 (January 1, 2017): 61–66. https://doi.org/10.1016/J.PROCIR.2017.03.130

“Review: On the History of the Automobile in Everyday Life on JSTOR.” Accessed July 3, 2022. https://www.jstor.org/stable/20081778

Ridzuan, Fakhitah, and Wan Mohd Nazmee Wan Zainon. “Diagnostic Analysis for Outlier Detection in Big Data Analytics.” *Procedia Computer Science* 197 (2021): 685–92. https://doi.org/10.1016/J.PROCS.2021.12.189

“Root Cause Analysis | Download Scientific Diagram.” Accessed June 15, 2022. https://www.researchgate.net/figure/Root-cause-analysis_fig7_264836341

Russell, Daniel M, Mark J Stefii, Peter Pirolli, and Stuart K Card. “The Cost Structure of Sensemaking.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’93*, n.d. https://doi.org/10.1145/169059

Rytkönen, Eelis, and Suvi Nenonen. “The Business Model Canvas in University Campus Management.” *Intelligent Buildings International* 6, no. 3 (2014): 138–54. https://doi.org/10.1080/17508975.2013.807768

S, Russell, and Bohannon J. “Artificial Intelligence. Fears of an AI Pioneer.” *Science (New York, N.Y.)* 349, no. 6245 (July 1, 2015): 252–252. https://doi.org/10.1126/SCIENCE.349.6245.252

Sagiroglu, Seref, and Duygu Sinanc. “Big Data: A Review.” *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems, CTS 2013*, 2013, 42–47. https://doi.org/10.1109/CTS.2013.6567202

Sahoo, Madhumita. “Evaluation of Machine Learning-Based Modeling Approaches in Groundwater Quantity and Quality Prediction.” *Advances in Remediation Techniques for Polluted Soils and Groundwater*, January 1, 2022, 87–103. https://doi.org/10.1016/B978-0-12-823830-1.00016-X

Sarker, Iqbal H. “Machine Learning: Algorithms, Real-World Applications and Research Directions.” *SN Computer Science 2021 2:3* 2, no. 3 (March 22, 2021): 1–21. https://doi.org/10.1007/S42979-021-00592-X

Sharma, Rashmi, Manish Prateek, and Ashok K. Sinha. “Use of Reinforcement Learning as a Challenge: A Review.” *International Journal of Computer Applications* 69, no. 22 (May 31, 2013): 28–34. https://doi.org/10.5120/12105-8332

Shmueli, Galit, and Otto R. Koppius. “Predictive Analytics in Information Systems Research.” *MIS Quarterly: Management Information Systems* 35, no. 3 (2011): 553–72. https://doi.org/10.2307/23042796

Singh, Amanpreet, Narina Thakur, and Aakanksha Sharma. “A Review of Supervised Machine Learning Algorithms.” *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, 2016, 1310–15

Sivarajah, Uthayasankar, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. “Critical Analysis of Big Data Challenges and Analytical Methods.” *Journal of Business Research* 70 (January 1, 2017): 263–86. https://doi.org/10.1016/J.JBUSRES.2016.08.001

Snyder, Hannah. “Literature Review as a Research Methodology: An Overview and Guidelines.” *Journal of Business Research* 104, no. March (2019): 333–39. https://doi.org/10.1016/j.jbusres.2019.07.039

Szepesvári, Csaba. “Algorithms for Reinforcement Learning.” *Synthesis Lectures on Artificial Intelligence and Machine Learning* 9 (July 6, 2010): 1–89. https://doi.org/10.2200/S00268ED1V01Y201005AIM009

Uyanık, Gülden Kaya, and Neşe Güler. “A Study on Multiple Linear Regression Analysis.” *Procedia - Social and Behavioral Sciences* 106 (December 2013): 234–40. https://doi.org/10.1016/J.SBSPRO.2013.12.027

Waller, Matthew A., and Stanley E. Fawcett. “Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management.” *Journal of Business Logistics* 34, no. 2 (June 1, 2013): 77–84. https://doi.org/10.1111/JBL.12010

Wang, Feng Hsu, and Hsiu Mei Shao. “Effective Personalized Recommendation Based on Time-Framed Navigation Clustering and Association Mining.” *Expert Systems with Applications* 27, no. 3 (October 1, 2004): 365–77. https://doi.org/10.1016/J.ESWA.2004.05.005

Wang, Huan, Min Ouyang, Qingyuan Meng, and Qian Kong. “A Traffic Data Collection and Analysis Method Based on Wireless Sensor Network.” *Eurasip Journal on Wireless Communications and Networking* 2020, no. 1 (2020). https://doi.org/10.1186/s13638-019-1628-5

Wang, Pei. “On Defining Artificial Intelligence.” *Journal of Artificial General Intelligence* 10, no. 2 (2019): 1–37. https://doi.org/10.2478/jagi-2019-0002

Ward, M. F. “Political Economy, Industrial Location and the European Motor Car Industry in the Postwar Period.” *Https://Doi.Org/10.1080/09595238200185441* 16, no. 6 (December 1, 2007): 443–53. https://doi.org/10.1080/09595238200185441

Williams, Graham. “Descriptive and Predictive Analytics.” *Data Mining with Rattle and R*, 2011, 171–77. https://doi.org/10.1007/978-1-4419-9890-3_8

Xiao, Yu, and Maria Watson. “Guidance on Conducting a Systematic Literature Review.” *Journal of Planning Education and Research* 39, no. 1 (2019): 93–112. https://doi.org/10.1177/0739456X17723971

Zott, Christoph, Raphael Amit, and Lorenzo Massa. “The Business Model: Recent Developments and Future Research:” *Https://Doi.Org/10.1177/0149206311406265* 37, no. 4 (May 2, 2011): 1019–42. https://doi.org/10.1177/0149206311406265

Internet sources

Ajitha, P V, and Ankita Nagra.: An Overview of Artificial Intelligence in Automobile Industry-A Case Study on Tesla Cars, < www.solidstatetechnology.us> (no given date), accessed March 2, 2022

Bush, Vannevar.: As We May Think, < http://www2.theatlantic.com/> (no date given), accessed July 3, 2022

David Ulrich, Wayne Brockbank: The HR Value Proposition, <

https://books.google.de/books?hl=en&lr=&id=iemTssZgYlwC&oi=fnd&pg=PA21&dq=value+proposition+&ots=hp_n_t2_S4&sig=wS_pL9ZuytC0PZhVjycfJ9bG7W8&redir_esc=y#v=onepage&q=value proposition&f=false> (no date given), accessed February 28, 2022

https://bmcintroduction.wordpress.com/channels/, accessed July 3, 2022

https://books.google.de/books?hl=en&lr=&id=pvo1SauGirsC&oi=fnd&pg=PA1149&dq=research+design&ots=qua-IBv-i7&sig=iuixC1jllEZs8gOIxEcioxV2q0c&redir_esc=y#v=onepage&q=research design&f=false, accessed June 28, 2022

https://drive.google.com/file/d/1_Ow8pxdXr2-u6KrrKphEG2UkeTAVYXOj/view, accessed February 28, 2022

https://www.proquest.com/openview/943bd867a26bf544f7c2212f81314ee0/1?pq-origsite=gscholar&cbl=35418, accessed July 3, 2022

https://www.statista.com/statistics/871513/worldwide-data-created/, accessed June 14, 2022

K.N.·奥托 (美), Kevin N. Otto: Product Design: Techniques in Reverse Engineering and New Product Development, < https://books.google.de/books?hl=en&lr=&id=0X54fSKq7bkC&oi=fnd&pg=PR13&dq=business+design+techniques&ots=54s9fDMpQ1&sig=cQrJrYNngWfAAXMtn3BUKej-j-w&redir_esc=y#v=onepage&q=business design techniques&f=false> (no date given), accessed February 28, 2022

Kamolson, Suphat: Fundamentals of Quantitative Research Suphat Sukamolson, Ph.D. Language Institute Chulalongkorn University, *Language Institute*, 2007, < http://www.culi.chula.ac.th/e-Journal/bod/Suphat Sukamolson.pdf%5Cnhttp://isites.harvard.edu/fs/docs/icb.topic1463827.files/2007_Sukamolson_Fundamentals of Quantitative Research.pdf> (no date given), accessed July 25, 2022

Martínez-Plumed, Fernando, Bao Sheng Loe, Peter Flach, Seán Seán´ Seán´o, ´ Eigeartaigh, Karina Vold, and José Hernández-Orallo.: The Facets of Artificial Intelligence: A Framework to Track the Evolution of AI, <

http://www.scimagojr.com/shapeofscience/> (no date given), accessed July 3, 2022

Porter, Michael E, James C Collins, Jerry I Porras, Mark W Johnson, Clayton M Christensen, Henning Kagermann, W Chan Kim, et al.: Reinventing Your Business Model, 2000 < www.hbr.org> accessed on July 5, 2022

Sheda, Sheda: How To: Business Model Canvas, < https://medium.com/seed-digital/how-to-business-model-canvas-explained-ad3676b6fe4a> (no date given), accessed June 28, 2022

Sukanya Bag, Analytics Vidhya: Which Machine Learning Algorithm Should You Use By Problem Type?, < https://medium.com/analytics-vidhya/which-machine-learning-algorithm-should-you-use-by-problem-type-a53967326566> (no date given), accessed June 21, 2022

Sykes, Alan O.: Chicago Unbound An Introduction to Regression Analysis, < https://chicagounbound.uchicago.edu/law_and_economics> (no date given), accessed June 24, 2022

Wesley Chun: Core Python Programming, < https://books.google.de/books?hl=en&lr=&id=mh0bU6NXrBgC&oi=fnd&pg=PR1&dq=what+are+features+of+python+programming+language&ots=XBjGyu5Zjb&sig=lKt94lX2d0NcuXodYMU_j5hX2S0&redir_esc=y#v=onepage&q=what are features of python programming language&f=false> (no date given), accessed June 22, 2022

DECLARATION IN LIEU OF OATH

I hereby declare that I produced the submitted paper with no assistance from any other party and without the use of any unauthorized aids and, in particular, that I have marked as quotations all passages which are reproduced verbatim or near verbatim from publications. Also, I declare that the submitted print version of this thesis is identical to its digital version. Further, I declare that this thesis has never been submitted before to any examination board in either its present form or in any other similar version. I herewith agree/disagree that this thesis may be published. I herewith consent that this thesis may be uploaded to the server of external contractors for the purpose of submitting it to the contractors' plagiarism detection systems. Uploading this thesis for the purpose of submitting it to plagiarism detection systems is not a form of publication.

Date: 16/08/2022 Canny Sen