**Reflective essay**

This project is based on a practical example of customer segmentation which is inspired by the approach of one of the biggest technology giants of the world like Amazon, Netflix and many other e-commerce and marketing giants. This approach has been effective since the time it was introduced and has been growing tremendously under various names but most popularly under the name of “Recommendation Systems”. This project aims towards the practical implementation of the customer segmentation technology which helps the business to better target their customers for increased customer acquisition and generating more revenue.

**Approach:**

This project lays out four ground domains or sections under which this entire project can be divided. These are frontend, backend, database and finally the machine learning part of the project (Morton, M, 2017) Deciding the technical stacks for these projects was very important because based on the selection of these technical stacks, the problems that were going to arise were supposed to be solved. Upon consideration, for the frontend I chose Angular is as its modularity and the advantage of dividing the entire code into components makes it efficient as a developer to build projects. For the backend, I chose Flask, because flask being a micro web framework and also based on python makes it easier to handle and develop backend applications as it has a huge list of libraries. I have choosen MongoDB as the NoSQL database since it has a good developer community and finally for the customer segmentation we selected K-Means Clustering as it was an unsupervised machine learning problem and K-Means is one of the most efficient methods when it comes to clustering (IBM Cloud Education , 2019).Now this entire project was to be hosted on cloud, so upon a lot of consideration I chose to go with Heroku Cloud Platform as this platform provides the simplicity and the ease of hosting due to the availability of the resources to solve the problems and resolve the doubts.

**Future Work**

Customer segmentation is the method involved with gathering clients in view of normal attributes. These customer clusters are useful in showcasing efforts, in recognizing possibly beneficial customers, and in creating customer dedication. The created model right now is efficient enough to handle any size of data and still give good results but what it lacks is the variety and the diversity of the data. Currently, the dataset that has been chosen has a limited features but this is quite evident too as we are currently using all free solutions and we do not have access to company size data which makes this project all the more important. The entire project is hosted on cloud platform with free resources and hence the response time of the features that we have built is a little more since everything is hosted on most basic hardware configurations. But despite these factors, the web application performs significantly faster with a response time of only around less than 2 seconds that too for rendering of machine learning models, and instantly for normal web pages and other applications. The NoSQL database hosted on MongoDB is also something that can be improved since now the database is hosted on a basic free plan on MongoDB cloud Atlas which increases the response time of the web application.(MongoDB ,2019).

**Limitations and Practical Challenges**

There were a lot of problems while building this project. The entire project was divided into two phases: the planning and the development phase. In the planning phase there were a lot of conflicts as to what technology stack to select because every stack had its own problems and advantages and to select the best one was difficult. After considerations I chose Angular for frontend and Flask for backend. Database selection was also going to be very difficult. Based on the dataset that I had, I could go for both SQL and NoSQL databases, but where SQL was easy to use and implement in local system, it was difficult to implement on cloud . Also out of all the NoSQL databases, MongoDB has the best resources and ease of use along with its own cloud deployment making it hassle free for us. As for the machine learning model that we have developed, that works amazingly well on the dataset that we have used but as we all know that that machine learning models are susceptible to even very small changes in the dataset. Since right now we are working a dataset with limited features, later while adding any more features, the model with its fixed parameters right now might not perform that well and the entire process of configuring the model parameters might have to be done again. There were also a lot of problems with the dataset that we have used specially related to the date columns. This issue was related to the date format in the csv file. While accessing the csv file, excel automatically detects the data column as date object type but in our case, it was not able to detect the date column on its own because of which while reading the date column, there were a lot of discrepancies and caused a lot of issues while performing operations int the code. To resolve this, I had to write a code to tackle this logic and convert each date element and format it to get the desired format and thereafter convert it into the date object which solved the error for us.

**Legal, ethical and Environmental Considerations**

The principles major aspects like include integrity, honesty, objectivity, responsibility, trustworthiness, impartiality, non-discrimination, accountability, robustness, usability, efficiency, and independence (Gupta, H., 2020). Utilising numerous statistical metrics and prediction methodologies, data science is valuable in identifying patterns and delivering efficient decisions.Segmentation is one of the best methods for analysing the motives and actions of consumers. Therefore, the goal of this study is to group ethical customers according to variables related to consumer ethics (i.e. actively benefiting, passively benefiting, questionable behaviour, no-harm, recycling and doing good).Data-driven segmentation enables businesses to build stronger relationships with their clients. They can use information to design new features, make informed judgments about retention, and strategically place their product in the market.

The data that is used in this paper are generic data with does not harm ethical values legally or environmentally.We created this using unsupervised learning, a kind of machine learning. In particular, we applied the K-means clustering clustering technique. After performing an analysis and data visualisation, we implemented our method.

**Personal Development**

It is a reality that education is not just about passing exams; it also involves comprehending the material and determining how what is learned in the classroom may be applied to real-world situations (Bolton, G, 2010). As a result, I've discovered that conducting extensive research and experimentation is necessary in order to apply the knowledge acquired in the classroom to related subjects and to apply the knowledge learned from books to real-world situations. Second, I've discovered that the most important skill for accomplishing personal objectives in both education and other areas of life is time management. After some contemplation, I've come to the conclusion that my improved performance in extracurricular activities and overall academic ratings is due to my ability to manage my time well and adhere to stringent deadlines.I got exposure to an expert level of data science and learned a great deal about my subject, and it inspired an interest in research I did not have before.

**References:**

Bolton, G, 2010, “Reflective Practice: Writing and Professional Development”, SAGE Publications Ltd, Hoboken, NY.

IBM Cloud Education (2019). *nosql-databases*. [online] Ibm.com. Available at: https://www.ibm.com/cloud/learn/nosql-databases.

MongoDB (2019). *NoSQL Databases Explained*. [online] MongoDB. Available at: https://www.mongodb.com/nosql-explained.

Morton, M. (2017). *5 Phases of the Project Management Process | TeamGantt*. [online] Teamgantt.com. Available at: https://www.teamgantt.com/blog/5-crucial-project-management-phases.

Gupta, H. (2020). Ethical Consideration When Collecting Data. [online] What After College. Available at: https://whataftercollege.com/data-science/ethical-consideration-collecting-data/.

VanMSFT (n.d.). *Securing SQL Server - SQL Server*. [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/sql/relational-databases/security/securing-sql-server?view=sql-server-ver16.