Milestone-2 - Literature Review - Individual Report

In my examination of the research paper [Sun16], I explored the analysis of a comprehensive dataset of NYC taxis utilizing Big Data technologies. The study's emphasis on leveraging MapReduce and Hive for understanding taxi network patterns within the Hadoop cluster showcased an innovative approach. By integrating the dataset into the Hadoop Distributed File System (HDFS) and employing tools such as Hadoop MapReduce, Hive, HBase, Pig, and Spark, the research efficiently processed and visualized data relationships and patterns. The paper's unique perspective of treating each taxi as a moving sensor within the NYC taxi system, coupled with the integration of historical data and a smartphone application, highlights its potential for providing valuable services to taxi stakeholders, enhancing the overall efficiency and accessibility of the taxi network.

Furthermore, my analysis of the research paper [Yazici13] illuminated the utilization of a large dataset of taxi trips to analyze decision-making processes among New York City taxi drivers, with a specific focus on enhancing access and passenger satisfaction at John F. Kennedy (JFK) Airport. The study's meticulous use of logistic regression to model driver decision-making, utilizing latitude and longitude coordinates and timestamps, underscored its significance in addressing crucial industry challenges and stakeholders' concerns. The adaptable nature of this methodology, applicable to similar datasets in various locations, adds to its potential for facilitating strategic improvements in taxi services beyond the NYC context.

I then contributed to the conclusion part. In the conclusion provided, our team's aim to harness the power of data analytics in predicting taxi fares, as evident in major industry players like Uber and Lyft, signifies our commitment to delivering a predictive solution that enhances transparency and user experience. By implementing comprehensive feature engineering on NYC-TLC's dataset and deploying a user-friendly prediction model, our project seeks to empower users with reliable fare estimates, aligning with the user-centric approach adopted by popular ride-sharing services. Our focus on leveraging the insights derived from data analytics underlines our dedication to enhancing the overall efficiency and accessibility of taxi services for users.