Introduction to Big Data & Data Science

OPTIMISING HEALTHCARE INSURANCE

Team Members

1. Alekya Mullapudi (11660468)
2. Siddhartha Goud Alwala (11661116)
3. Jeevan Parakala (11706336)
4. SaiTeja Goud Akkala (11703752)

Project Description:

In today's competitive health insurance industry, maximizing revenue while delivering value to customers is a constant challenge. Our project, "OPTIMISING HEALTHCARE INSURANCE" aims to empower a leading healthcare insurance company to thrive in this dynamic landscape by harnessing the power of Big Data.

Input Data:

- Data Sources: Data from various sources, including customer information, insurance policy details, and competitor data. This data can include structured and unstructured data such as text, numerical values, etc.
- Raw Data: Raw data from these sources may be in different formats and need preprocessing

Output Data:

- Insights and Reports: Data-driven insights and reports are generated from the analysis of the stakeholders within the company. These reports may include customer behavior trends, competitor insights, and revenue enhancement strategies.
- Data Pipelines: Developed data pipelines for ongoing data processing and analysis.

Motivation:

In the ever-changing healthcare insurance industry, it is becoming increasingly hard to remain competitive and profitable. Our project, "OPTIMISING HEALTH INSURANCE" aims to provide a powerful solution to a leading health insurance company.

- Revenue Enhancement: The motivation for this project is to allow the health insurance funds to improve its revenue. In a highly competitive market, increasing revenue is not the only goal. It is also essential for sustainability and economic growth.
- Understanding the Customer: Understanding the complexity of customer behavior, preferences & needs is of utmost importance in the insurance industry. The project aims to shed unprecedented light on these aspects. By analyzing massive amounts of data from various sources, including web scraping and third-party data providers, we aim to provide a company with a comprehensive view of its customer base.
- Personalized Offers: The motivation of the project extends to creating a more individualized and responsive insurance experience for customers. Using the insights gained from data analysis, the company can tailor its offers to individual customer profiles. It also increases the value proposition to customers and strengthens customer loyalty.

- Customer Loyalty: Recognizing and rewarding loyal customers is the key motivation for this project. By developing algorithms to calculate royalties or rewards for customers who consistently purchase policies, the company can strengthen customer loyalty. This approach not only boosts customer retention but also helps to increase revenues over the long term.

By harnessing the power of big data, advanced analytics, and innovative strategies, we aim to position the company for sustainable success in the health insurance industry. This project underscores our commitment to data excellence and providing solutions that enable our clients to succeed in a competitive market.

Significance:

- Customers can benefit from personalized insurance offers that match their specific needs and preferences. It means they are more likely to get coverage that suits them, potentially saving them money and providing better protection.
- Improve Customer Experience: With a deeper understanding of customer behavior, the company can enhance the overall customer experience. It might include smoother claims processing, faster responses to inquiries, and more accessible support.
- Better Health Outcomes: An analysis of customer data can lead to the development of wellness programs or incentives for healthy behaviors, which could encourage customers to adopt healthier lifestyles, leading to improved health outcomes.

Objectives:

To establish a comprehensive Big Data infrastructure tailored for a healthcare insurance firm with the dual objectives of revenue enhancement and improved customer comprehension. This framework will conduct a

recurrent analysis of competitor data derived from diverse origins, encompassing web scraping and third-party sources, to monitor customer behavior and well-being. The insights garnered will be utilized to craft personalized customer offers and ascertain loyalty-based royalties. The overarching aim is to equip the company with the tools and insights necessary for making well-informed business decisions that result in increased revenue.

Specific goals include:

- Form a data pipeline for collecting, cleansing, and processing of data from the datasets.
- Construct a robust Big Data ecosystem for efficient storage and analyzing the amassed dataset.
- Develop machine learning models proficient in predicting customer behavior and well-being.
- Crafting algorithms adept at tailoring offers to individual customers and determining royalties for those with a history of loyalty.

FEATURES:

The project deliverables include comprehensive datasets capturing customer behaviors and insurance variables. Cleaned, analysis-ready data, an organized database system, a data transfer system using Sqoop, efficient data analysis with Hive, Spark SQL integration for enhanced querying, and data visualizations for easy comprehension. The project's uniqueness lies in its emphasis on highly personalized insurance offers, real-time insights for agile decision-making, and the creation of a robust data analysis ecosystem by combining Apache Spark, Hadoop, Sqoop, and other tools.

VISUALIZATION:

- After transferring the data into HDFS, we established a connection with Spark. Subsequently, we employed Python for data analysis. During this phase, we obtained the desired results in a tabular format.
- The results of each use case can also be shown as graphs to make better business decisions.

REFERENCES:

- Books and Reports:

i. Beranger, Jérôme. 2016. Ethics in Big Data: the medical datasphere. London: Elsevier.

ii. Davis, Cord and Patterson, Doug. 2012. Ethics of Big Data. Farnham, O'Reilly.

- Web Sites:

i. Big Data Ethics

ii. GeekforGeeks