Capstone #3 Project Ideas

1/ Loan default classification

The finance sector is focused around one essential mathematical problem – how can we assess and quantify risk? While this is usually calculated by large firms, in recent years more and more opportunities have arisen for individuals to not only buy but also sell financial products. LendingClub, a financial services company headquartered in San Francisco, California, enables borrowers to create unsecured personal loans between $1,000 and $40,000 and investors to search and browse the loan listing on LendingClub website. This puts normal people in the same position as banks, allowing them to select loans that they want to invest in based on the information supplied about the borrower, amount of loan, loan grade, and loan purpose.

But with individuals acting as banks, they then have the same problem as banks do – how can they accurately assess the risk of giving a loan to maximize their return? With Machine Learning, I aim to help answer this question by building a model that can evaluate and learn from previous loans to help recommend the best loans for individuals to invest in.

Data link: <https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv>

2/ Loan default classification big data

The purpose of this capstone is to learn how to work with big data. Thanks to data science, massive data sets especially in finance, healthcare, and online services can be used to address business problems that would not have been able to be tackled before. I will use Pyspark in Databricks environment to build a machine learning model to predict a larger loan default dataset. Data link: <https://www.kaggle.com/datasets/ethon0426/lending-club-20072020q1>

3/ Time series to predict stock prices

I always liked finance and am fascinated by the moving of stock price over time. I would like to build a time series model to track the performance of S&P 500 which the is most famous financial benchmark in the world and identify the strong and weak stocks in the list. Data link: https://www.kaggle.com/datasets/andrewmvd/sp-500-stocks