SNEHA KANNAN

I am adept at processing large quantities of complex data, feature engineering, building and deploying machine learning models in production by collaborating with product managers, data engineers and software developers. I am seeking a new opportunity to utilize my skills and broaden my experience while I contribute to the growth of the organization.

WORK EXPERIENCE

2020 Mar

Data Scientist II

Braviant Holdings — Chicago, IL

- Analyzed employer names from the loan applications by fuzzy matching and categorized them into common industry types to understand the impact of the current economic scenario on the product

2020 Feb | 2017 July

Senior Analyst

Portfolio & Fraud Analytics, Enova International — Chicago, IL

- Reduced non-payment rate by 300 bps with a potential saving of 800K R$/year by building a GBM model using transactions from Bank Statements and deploying it in AWS Sagemaker
- Analyzed the profitability of funded loans and helped determine budgeting targets for the portfolio by assessing the Net Present Value incorporating marketing costs, underwriting and operational costs
- Reduced funding fraud loan applications and potentially saved 1.1M R$/year by building a fraud tracking database with advanced reporting and real time dashboard using web app framework Flask Python
- Increased productivity for fraud operations team by 25% using a recommendation model to generate a similarity score between all loan applications and known fraud loans
- Conducted various ad-hoc analysis by using complex PostgreSQL queries and statistical techniques to improve on the existing risk management, marketing and operational strategies

2017 June | 2015 July

Analyst

Portfolio Analytics, Enova International — Chicago, IL

- Increased the recovery rate from defaulted loans by 11% by developing a collections model to predict the likelihood of payment and optimized the debt collections process
- Improved conversions and reduced default rate by building a model suite to carry out A/B tests during loan processing while reducing risk to the product and lowering friction to the end users
- Reduced the underwriting cost by 250K R$/year by optimizing the underwriting process with the help of a machine learning model and ensured better customer segmentation
- Built a statistical model facilitating the direct mail campaign for a product which provided a 23% lift in KS and increased the AUC by 0.04 compared to the existing model

CONTACT

Website
Github
Linkedin

SKILLS

Programming: R, Python, PostgreSQL, Java, GitHub

Tools: tidyverse, caret, ggplot2, tidymodels, pandas, scikit-learn, NumPy, Jupyter Notebook, Microstrategy

Machine Learning: Regression & Classification models (Linear & Logistic Regression, Random Forest, Boosting, Support Vector Machines), Model Selection and Evaluation, Hypothesis testing (A/B testing), Clustering, Segmentation, Optimization, MapReduce

Made with the R package pagedown.

The source code is available at github.com/snehakannan/cv.

Last updated on 2020-04-01.

EDUCATION

2015 May | 2013 Aug

Master's in Management Information Systems
University of Illinois at Chicago — Chicago, IL

GRADUATE PROJECTS

Exploring jobs on Analytics: Natural Language Processing (NLP)

- Identified different features of jobs related to Analytics & Big data by collecting data from a popular job listings website. Discovered word clusters using the technique of topic modeling on the job descriptions

Item-based collaborative filtering for movie recommendation

- Developed a MapReduce program in Hadoop to compute Pearson correlation similarity for all pairs of movies in the data set. Used this to compute the prediction on a movie for a given user using the weighted sum technique and recommended top k movies to the user

Frequent itemset mining using MapReduce

- Generated 2 and 3-item sets using MapReduce from a large transaction data set using association rules. Identified items which were frequently bought together and those which were bought independently

Clustering using K-means under Apache Mahout

- Preprocessed and cleaned the JSON input files of the Yelp reviews data set. Applied K-means clustering under Mahout and Hadoop environment to find clusters in the reviews

Strategic Operating Plan to a group of Department stores

- Used the SQL Server Analysis Services Cubes for analyzing the stores data available through the University of Arkansas website. Created a report which identified the stores which were least profitable using the profitability KPI metrics, days on which the stores were least profitable so they can reduce the operations on those days and least profitable brands in the stores so their inventory can be reduced