Sanghamesh Shiddalingesh Vastrad

Data Scientist | sanghamesh@cs.toronto.edu | (647) 580-6884 | Toronto sanghamesh.netlify.app

EXPERIENCE

ICICI Bank, *Toronto, Canada* *8 months*
Machine Learning Applied Research Intern | May 2020 – Present (Dec 2020)

- Sole member of the Data Science and Analytics Group at ICICI Bank Canada with research approved and partly funded by Mitacs Accelerate. Trusted with designing and building a 'Zero Credit Touch' (ZCT) system with projected impact on 20 million+ customers.
- Improved credit underwriting gini coefficient (2*AUC - 1) by 12% and decile wise gini by 6% by developing an ensemble model consisting of XGBoost, LightGBM and a neural network (with LSTMs and 1D Convolutions).
- Cut down data dimensionality by 55% from 400 to 185 features using Shapley model explanations and manifold learning techniques like Locally Linear Embeddings and t-SNE.
- Optimized AUC using synthetic data generation using a combination of SMOTE and CTGAN (Conditional GAN for Tabular Data) after researching best-suited techniques for handling imbalanced datasets.
- Used Latent Dirichlet allocation (LDA) for customer segmentation to identify clusters where the model does not perform well, that translated to an increase in model's decile wise gini by 11% when the poor performing cluster is removed.

Western Digital (SanDisk), *Bangalore, India* *1 year 6 months*
Software Developer | Jul 2018 – Jul 2019

- Strongly appreciated by senior management for delivering key features for managing, verifying, and visualizing iNAND firmware algorithms using C# and XAML in an agile environment meeting over 90% deadlines.
- Spearheaded prototyping for 'image-save' capability for SSDs of 2TB+ capacity using Spark and Kafka to make data processing 400% faster. Led data visualization tasks using Live Charts for Windows Presentation Foundation (WPF).

Software Engineering Intern | Jan 2018 – Jun 2018

- Reduced version control testing time from 72 to 8 hours by designing and implementing an automation framework using Python.
- Presented over 20 data analysis results and reports of multiple HDD performance benchmarking programs.

EDUCATION

MSc in Applied Computing (Data Science Concentration) Sep 2019 – Dec 2020
University of Toronto, Dept. of Computer Science GPA: 3.93/4

- Coursework in Data Science, Machine Learning, Computational Social Science and Human Computer Interaction.
- Graduate Teaching Assistant for Software Design course.

BEng in Computer Science & Engineering Aug 2014 – May 2018
Sri Jayachamarajendra College of Engineering, India Rank 2 (of 150+) GPA: 9.81/10

- Runner-up, 'Code for Good' India hackathon, J.P. Morgan Chase & Co., 2017.
- Received Huawei Technologies Scholarship twice for University Top 10 (out of 800+) in 2016 and 2018.

TECHNICAL SKILLS

Data Science & Machine Learning	**Python** - NumPy, Scikit-learn, Pandas, xgboost, lightgbm, imblearn, StatsModels **R** - tidyverse (dplyr, ggplot2, broom)
Data Visualization	Plotly, matplotlib, seaborn, ipywidgets, Voila, streamlit
Deep Learning	TensorFlow, Keras
Big Data Engineering	BigQuery, Hadoop, Spark
Application Development	Java, C#, C, JavaScript, SQL

PROJECTS

Image-based Plant Disease Detection App using MobileNet | Nov 2017 – May 2018

- Achieved 95.7% accuracy for detecting plant disease in real time, based on leaf images by training a MobileNet using transfer learning (TensorFlow Mobile) for 15 species of plants and 38 types of diseases.
- Developed 5+ features for description and remedies, consultation with a plant pathologist, and geotagging (in Android).

Accuracy, Recall or Explainability? - A Case Study in Predicting Readmission for Diabetes | Jan 2020 – Feb 2020

- Explored tradeoff between metrics like ROC AUC, Accuracy and Recall for short-term and long-term risk prediction. Overcame data imbalance through class weight hyperparameter in XGBoost to increase Recall to 0.65.
- Used SHAP (Shapley values) for feature importance and contribution dependence plots to make suggestions for reducing readmission.

Averting Algorithm Aversion through Explainability | Jan 2020 – Apr 2020

- Conducted an online lab (A/B testing) experiment through Qualtrics Survey Platform and Amazon MTurk to determine the effect of model explainability on machine learning model aversion. With empirical proof, showed that explainability has a small effect (Cohen's d=0.04) on how people rate the algorithm and improves the chance of choosing algorithmic forecasters being chosen.