SANKIRNA JOSHI

513-410-8702 | joshis2@mail.uc.edu | linkedin.com/in/sankirna | sankirnajoshi.github.io

SUMMARY

Data Scientist with 5 years of industry experience. Proven track record of developing and delivering machine learning models and data-driven solutions from conception through release. Excited and passionate about using my skills to solve challenging problems by building products that leverage data-driven techniques.

SKILLS

- Languages: Python | R | SQL | Stan | Bash | Html
- Data science tools: Pandas | Numpy | Scikit-learn | matplotlib | plotly | tensorflow | transformers | rstan | tidyverse | ggplot2
- Machine learning: Statistical modeling | Bayesian statistics | XGBoost | SVM | KNN | Trees and Random Forests | Regression methods | Neural networks | A/B testing | Hypothesis testing | NLP

PROFESSIONAL EXPERIENCE

Data Scientist, Afiniti (Washington DC) **Sep 2020 - Present**

- Working on building, testing, deploying, and monitoring AI algorithms for optimizing customer experience and employee productivity through intelligent call routing for large telecommunication businesses. Major clients include Verizon and Comcast.
- Developing and productionizing Bayesian IRT models for identifying patterns and interactions between agents and callers to drive profitable outcomes.
- Developed predictive models on datasets of over 100 million customers and thousands of attributes by using advanced statistical modeling, machine learning and data mining techniques. Developed new features and scores using the model outputs as feeds to a downstream Bayesian model.
- Built a proof-of-concept model using unsupervised techniques to demonstrate that groups formed within the data, possess discriminating powers to separate data on key design metrics and thus aid in stronger machine learning capabilities.

Software Engineer - Data Science, InCred **Aug 2017 - Apr 2019**

- Designed and owned the data warehouse infrastructure; built ETL pipelines using Python and SQL to load and process a variety of data types to build the data science infrastructure at the company.
- Created the company's first data pipelines and statistical models with focus on automation and fraud detection with sensors in place to run the pipelines at distinct steps in an application's lifecycle.
- Led the production of exco-level dashboarding on company loan portfolio, credit risk metric, sales insights, customer delinquency and ad-hoc reports.
- Applied natural language processing techniques to detect duplicate applications in the system. Developed a POC spelling mistake identification tool using a custom dictionary to validate hand filled inputs and suggest possible corrections.

Data Analyst, Cognizant **Aug 2015 - Jul 2017**

- Integrated data from multiple sources using Python to feed into statistical and visualization reports that measured the performance across different businesses for our client.
- Utilized ETL and SQL for data wrangling, data cleaning, and data manipulation and loaded transformed data into data warehouse.
- Developed ETLs to capture the slowly changing nature of data (SCD II) and load into Teradata enterprise data warehouse using Informatica Workflows and Teradata BTEQ scripts.

EDUCATION

Masters Business Analytics, University of Cincinnati
Machine Learning Engineer Nanodegree, Udacity

DATA SCIENCE PROJECTS (below and more @ sankirna.joshi@github.io)

- **Web application for sentiment analysis** (Python, Dash, AWS)**:** Github
  Developed and deployed an LSTM model in a Dash framework and deployed using AWS. Achieved 65% accuracy on 5 classes
- **Carvana Image Segmentation Challenge on Kaggle** (Jupyter notebooks, Google Cloud)**:** Github
  Built a 100-layer deep NN model to remove the backgrounds from car images and achieved a dice accuracy score of 0.9944.
- **Cincinnati Crime Exploration 2019** (Tableau Public): Tableau
  Developed an exploratory dashboard for visualizing crime statistics in the Cincy suburbs.
- **Dog Classification using Transfer Learning** (Python, Keras, Jupyter Notebooks): Github
  Used Inception model to apply transfer learning for classification of 133 dog breeds. Achieved a test score accuracy of 84%