ANKIT AGRAWAL
Data Scientist | Kaggle Expert

(801)-800-1304 | kit.agrawal1@gmail.com | https://www.linkedin.com/in/ankitagrawal3/

Business-minded data scientist with demonstrated ability to deliver valuable insights via data analytics and provide data-driven solutions to business problems.

SKILLS

Languages: Python, R, C++, SQL, Matlab, HTML, CSS

Data Science: Data Analysis, Predictive Modeling, Big Data Queries, Data Mining, Visualization Tools, Machine Learning Algorithms, Forecasting, Natural Language Processing, Statistics, A/B testing, Linear Algebra.

Packages & Tools: pandas, scikit-learn, tensorflow, pytorch, opencv, matplotlib, nltk, seaborn, Facebook Prophet, Hugging Face, beautifulsoup, Tableau, Github, Docker, Google Cloud Platform, AWS SageMaker, AutoML.

EXPERIENCE

Aakash 88 LLC — The Woodlands, TX
Data Scientist — November 2019 - October 2021

Performed data analysis using Tableau, unsupervised learning methods and feature engineering to gain insights in the data. Built and maintained deep learning pipeline to forecast hourly electricity price & load for wind generated electricity trading for ERCOT market.
- 13% increase in annual profits.
- 35% reduction in analysis time through automation.

The University of Utah — Salt lake City, UT
Machine Learning Researcher — January 2017 - August 2019

Run experiments on cluster (CHPC) to detect optimal subspace configuration for SMACK parameters out of 5 Billion possible configurations.
- Reduce TIMEOUT errors by 23% across 14 benchmark categories.
- Reduce Type II error by 3% across 14 categories by using AutoML and ETL data pipelines.

PROJECTS

SV-COMP program classification: Given 10,000 C-programs (or 10 million lines of code), performed feature extraction, feature engineering, and implemented *classification trees* based models to classify them into appropriate SV-COMP benchmark categories.
- *98.3% classification accuracy* compared to previous 67% accuracy.
- *3x speedup* than previous implementation.

Google landmark detection: Given 6 million train images and 200,000 test images, implemented 2-layer image classification model firstly to classify whether an image contains a landmark using *VGG-19 transfer learning* by injecting non-landmark images in the training dataset and secondly, predict appropriate landmarks in the images containing landmarks using *ResNet transfer learning*.

Self-driving car: Given a train video with 25,000 frames and speed of car at each frame, implemented a *CNN - RNN regression model* to predict speed of the car for a test video with 8,000 frames.
- Achieved MSE < 10.

PAPERS

SMACK software verifier and verification toolchain, Z. Rakamaric et.al

EDUCATION

The University of Utah — Salt Lake City, UT
PhD Candidate, Machine Learning — August 2019
MS, Computer Science — December 2016