

Ankit Agrawal

Data Scientist | Kaggle Expert | Mentor
kit.agrawal1@gmail.com | kitagrawal.github.io

TECHNICAL SKILLS

Languages: Python, R, C/ C++, SQL, 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, beautiful soup, Tableau, GitHub, Docker, Google Cloud Platform, AWS Sage Maker, AutoML.

EXPERIENCE

Data Science Mentor

Great Learning

December 2021 - Present

Topics covered include Statistics, Random Forest, SVM, Linear and Logistic regression, KNN, ANN, CNN, LSTM, PCA, LDA, Spectral clustering, K-means, GMM, Autoencoders, Recommender systems, Graphical models (HMM, SI, SIR), Feature engineering tools (Deep feature synthesis, AutoML).

- Mentored over 60 students with average rating of 4.6

Data Scientist

Aakash 88 LLC

The Woodlands, TX

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.

Machine Learning Researcher

The University of Utah

Salt Lake City, UT

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

Fraud detection for Ethereum Trading: Given Ethereum transaction details, successfully built a classification model to detect fraudulent Ethereum transactions.

- Handled problems like missing data, imbalanced data, duplicate transactions, verifying transaction contracts, outlier detection and handling, feature selection.

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.

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.

EDUCATION

The University of Utah

PhD Candidate, Machine Learning

MS, Computer Science

Salt Lake City, UT

December 2016