

Apoorva Shah

DATA SCIENTIST · MACHINE LEARNING ENGINEER

Los Angeles, CA

☎ (+1) 609-721-1347 | ✉ apoorvashah101@gmail.com | 📱 apoorvashah101 | 🌐 apoorvashah101

Education

University of Southern California

Los Angeles, CA

PH.D. IN MATHEMATICS (STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS)

August 2022

- **Awards:** Theodore Edward Harris Teaching Prize in Mathematics
- **GPA:** 3.91/4.00

University of Chicago

Chicago, IL

B.S. IN MATHEMATICS AND STATISTICS

June 2017

Skills

| | |
|-------------------------|---|
| Programming | Python, R, SQL, Haskell, Java, LaTeX |
| Machine Learning | Scikit-Learn, Tensorflow, Numpy, Pandas |
| Languages | Mandarin (Intermediate-Advanced) |

Work Experience

CDW Data Science and Analytics

Vernon Hills, IL

DATA SCIENCE CONTRACTOR

June 2021 - January 2022

- Created a product ranking methodology using linear regression in order to optimize customer Add-to-Cart rate.
- Presented and analyzed a list of proposed search-ranking algorithms for consideration along with industry use case examples.

DATA SCIENCE INTERN

- Forecasted weekly sales of the top 20 products and services over a period of 8 weeks using an LSTM Encoder-Decoder model.
- Implemented data preprocessing, outlier handling, and data smoothing options on the entire 4GB dataset (7 million data points).
- Streamlined thousands of separate time series forecasting models into a single multivariate model, thereby reducing training time.

Projects

IMDb Sentiment Analysis with Data Augmentation

NLP

- Performed a sentiment analysis in TensorFlow on the IMDb movie reviews in order to predict whether a movie review was positive or negative.
- Improved the accuracy of a pre-trained BERT model by applying Easy Data Augmentation (EDA) techniques.

Finding Donors for CharityML

SUPERVISED LEARNING

- Deployed supervised learning models in Python using the Scikit-Learn library to target potential donors for a hypothetical charity organization using data from the 1994 Census.
- Achieved an 85% accuracy at predicting which individuals earned over \$50k per year (potential donors), outperforming the benchmark of 25%.

1992 LA Riots Casualty Analysis

DATA VISUALIZATION

- Produced a geospatial analysis with Folium by clustering casualties during the 1992 Los Angeles riots and identified regions with high casualties.

Flower Image Classifier

DEEP LEARNING

- Built and trained an image classifier in Python for the Oxford Flowers Dataset using a pre-trained model from TensorFlow Hub.
- Created a python script to allow the user to input any flower image for classification.

Topic Modeling with Clinical Notes

NLP

- Clustered sample clinical notes by similarity using Latent Dirichlet Allocation (LDA) in order to understand distribution of topics amongst electronic medical records.
- Preprocessed the data using customized stopwords to account for the overall medical nature of the corpus.
- Determined the dominant topic for each document and displayed the most representative medical record for each topic.
- Calculated the optimal number of document groupings using coherence scores.