Zachary Apell

1820 Country Club Dr., Cherry Hill, NJ 08003
Phone: (856) 298-0064 E-mail: zachapell220@gmail.com
Linkedin: www.linkedin.com/in/zach-apell Personal Website: https://zapell.github.io

Skills

Computer: Proficient in Python, PySpark, SQL, R, C++, Experienced in SAS, Hadoop, STAN, ReactJS, Javascript, MongoDB
Python Libraries: Pandas, Numpy, Tensorflow, Scikit-learn, Keras, NLTK, PyTorch, Seaborn, Matplotlib
Tools: Amazon Sagemaker, Jupyter Notebook, AWS (Certified Cloud Practitioner), Tableau, Snowflake, Git, RStudio, Experienced with Excel

Education

University of Michigan, Ann Arbor, Michigan May 2020
Bachelor of Science in Data Science

- 3.25/4.0 GPA
- Coursework: Machine Learning, Data Mining, Applied Regression Analysis, Bayesian Data Analysis, Web Systems, Database Management Systems, Honors Multivariable Calculus, Computer Vision

Work Experience

Rocket Mortgage **Detroit, MI**
Data Scientist April 2022-Present

- Dial Prioritization Models - Likelihood of lead to close
  - Developed full Data Science workflow from raw data through deployment using Amazon Athena, Sagemaker, and S3 for purchase and refinance retention models
  - Leveraged PySpark to build dataset joining data from legacy data lake with new AWS data lake
  - Created preprocessors and unit testing for categorical and numerical values used in lambda handler function for inferencing
  - Created Hyperparameter tuning jobs to determine the best XGBoost model parameters
  - Implemented Platt Calibration for comparison of true probabilities between different models due to highly imbalanced datasets
  - Improved top 3 decile conversion performance by 75% compared to old standard
- Prioritization Engine Model (PEM) Pre-Credit Purchase Test
  - Partnered with Revenue Operations to identify opportunity for clients to be escalated to new banker
  - Designed query to recreate complex business logic to identify control group with 95% accuracy
  - Built XGBoost model to predict loans most likely to reach PAL status that converted at higher rate than control group
  - Improved folder conversion by 31% using escalation logic overall, top 2 deciles converted at 150% better rate

Freedom Mortgage **Mt. Laurel, NJ**
Data Scientist June 2020-March 2022

- Designed query to connect to Tableau in real time for a catalog of rule triggers to allow transparency for compliance, regulations, and operations business divisions
- Leveraged experience with Data Governance, ETL, and Marketing Analytics business groups to create SQL queries to join data from Marketing, Servicing, Originations, and Customer data sources for use in Lead Prioritization and Customer Satisfaction models

Wacker Chemical Corporation **Ann Arbor, MI**
Data Science Intern June 2019-August 2019

- Led Sales Forecasting project
  - Experimented with Long Short-Term Memory Recurrent Neural Networks, Support Vector Regression, and SARIMA models to forecast sales and order entries for business team using internal sales data and external economic factors
  - Achieved 87% accuracy on order entry forecast and 79% accuracy on sales forecast
  - Improved sales forecast by 10% less deviation in 2019
  - Prepared dashboards using Tableau for business teams by visualizing customer and product trends and insights
- Natural Language Processing Analysis
  - Used Twitter API to scrape tweets containing hashtags and mentions pertaining to specific customers and market segments
  - Used Microsoft Azure API and NLTK library to calculate sentiment analysis of tweets and extract key phrases from tweets
  - Presented to business team with insights about current and prospective customers to help sales and marketing teams

Personal Projects

Live Betting Database August 2020-January 2022

- Automated a process using Python and selenium remote webdriver to scrape dynamic data on an AWS EC2 instance to a PostgreSQL database hosted on an AWS RDS instance
- Utilized cronjobs and AWS cloudwatch to schedule bash scripts and lambda functions to run at specific times to minimize total cost

Spotify Analysis August 2019-January 2020

- Clustered saved songs based on features grabbed from Spotify API to determine similarly structured songs between genres
- Visualized the audio features by building a Tableau dashboard