

SASHA PROKHOROVA

DATA SCIENTIST AT METIS

SUMMARY

Engineer with strong linguistic aptitude is leveraging experience in Python and SQL to a full-time career in data science. Interested in the intersections of engineering and natural language processing.

CONTACT

✉ prokhorova.sasha8@gmail.com
🌐 tinyurl.com/y2dxu72z 📞 (831) 334-3482
in aaprokhorova/ 🗣️ sasha-talks-tech

EDUCATION

San Francisco State University
Bachelor of Science Electrical Engineering 2019
Focus on Embedded Systems and Robotic Control

Smolensk State University
Bachelor of Arts Linguistics

SKILLS

PROGRAMMING LANGUAGES: Python, SQL, HTML/CSS
LANGUAGES: English, Russian, Spanish, German
PYTHON PACKAGES: Numpy, Pandas, SciKit Learn/ Statsmodels, Beautiful Soup, NLTK, Gensim, SpaCy
CLOUD AND DATABASES: AWS (Elasticsearch, Sagemaker, Kibana), PostgreSQL
DATA VISUALIZATION: Tableau, PyLDAvis, Matplotlib, Seaborn
SUPERVISED LEARNING: Linear Regression, Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Naive Bayes, Decision Trees, Random Forest, Gradient Boosting, Time Series
UNSUPERVISED LEARNING: Clustering, Dimensionality Reduction, Natural Language Processing

VOLUNTEERING

Santa Cruz Homeless Shelter
Santa Cruz
Helped with meal serving and preparation. This was a community outreach attempt to raise awareness about humanitarian causes, such as poverty alleviation and destigmatization of mental health issues.

EMPLOYMENT

Metis
San Francisco, CA
Data Scientist · June 2020 to Sept. 2020

Metis is an ACCET accredited 12-week immersive data science bootcamp focuses on Python, statistics, supervised and unsupervised machine learning, data analysis, data visualizations, database management, and communication. Designed five end-to-end projects utilizing data acquisition, cleaning and modeling to derive data-driven insights.

Predicting Business Success

Built a classification model predicting restaurant star ratings based on foot traffic and attributes. KNN, SVM, Logistic Regression and Random Forest were utilized to determine the most impactful attribute.

Yelp Review Data

Performed NLP analysis and topic modeling on Yelp restaurant reviews, using LDA, TF-IDF vectorizer, gensim, NLTK, and SpaCy. Created an automated text pre-processing pipeline. Customer reviews were analyzed to determine what aspects of the dining experience are crucial for a restaurant success.

Women through the lens of The New York Times

Data journalism research on gender equality and representation in The New York Times. Data collection through the Archive API for 1950 - 2020, topic modeling using SpaCy, gensim and LDA. Frequent term visualization based on the keywords and the timeframe. Sentiment analysis using NLTK Vader.

Vimix AI Solutions San Francisco

Data Engineering Intern · Apr. 2019 to Nov. 2019

- Coordinated a webinar series for ML and AI solutions, increasing the number of companies interested in the product by 40%
- Generated content promoting brand recognition and awareness
- Created dynamic dashboards to aid companies in the AI decision-making
- Analyzed user-generated videos to predict product perception and marketing ROI, with predicted savings of 30% of marketing efforts per company
- Collaborated with engineers on performing data analysis in AWS Kibana using Elasticsearch, improving report delivery efficiency by 35%
- Organized a virtual conference for ML and AI solutions, attracting 2 new potential investors

Inland Metal Technologies Hayward, CA

Industrial Engineering Intern · Oct. 2017 to Feb. 2018

- Created and implemented a POC to compare Asana vs. Confluence project management tools, lowering software costs by 15%
- Worked with three engineering departments to recommend lower cost manufacturing options
- Updated process mappings for three robotic welding stations and assembly lines
- Performed time studies, updated time standards and ran cost analysis in Excel by calculating labor hours and the cost of each part produced, decreasing labor costs by 8%