Elda Pere
Data Scientist

San Francisco Bay Area | (973) 897-8567

Email: eldapere8@gmail.com
Portfolio: https://eldapere8.github.io/
LinkedIn: https://www.linkedin.com/in/elda-pere/

SKILLS

Languages & Software: Python (Pandas, Scikit-learn, Tensorflow, Plotly), Amazon Web Services, Google Cloud Platform (BigQuery, AI Platform), Java, R, SQL, Flask, Apache Kafka, Spark, Hadoop, Tableau, Google Data Studio, Google PLX

Statistics and Machine Learning: Probability & Causal Inference, Field Experiments, Natural Language Processing, Hypothesis Testing, Optimization, A/B Testing, Classification & Clustering, Regressions, Neural Networks, Deep Learning

WORK EXPERIENCE

Accenture, Advanced App Engineering Analyst   05/2021 – Present

FANG Company, Data Science Practitioner
- Developed forecasting models to predict product performance for Maps app marketing metrics. Reduced previous project lead time by half while improving accuracy by 40% with feature engineering
- Built prediction models using MLP regressor, random forest regressor, FB Prophet models at over 90% R^2
- Constructed an optimization model using a fitted cost function to allocate marketing budget across countries with respect to given constraints. Assisted with building the model-interface pipeline in GCP

FANG Company, Analytics Consultant
- Employed data analytics solutions to improve the efficiency of internal program training using Google PLX, Data Studio and SQL in an agile process
- Ensured validity of reported metrics for consumer-facing campaigns of 1M+ users by running data validation, auditing and hypothesis tests for new certification program

Happiness Alliance, Data Scientist
- Leverage data science to help the non-profit Happiness Alliance build new economic paradigms focused on individual happiness, community wellbeing and ecological sustainability
- Use Jupyter Notebooks and Tableau to set up a data pipeline and perform regression analysis on given data

Self-Employed, Data Science Consultant   06/2020 – 05/2021
- Deployed a logistic regression in AWS to predict whether voltage drops originated onsite or from the power grid for Powerside Inc. Creative feature engineering led to 98% accuracy and a reduction of $2,500/month in computing costs
- Increased donations and operation efficiency for Nehemiah Gateway USA by working on a marketing and fundraising analytics dossier including NLP text analysis on newsletters, social media visualizations and A/B testing for email blasts
- Transformed Le Gungh Hotel operations by building a MySQL database for their hotel and spa in Vietnam

ViviendasLeon, Data Scientist   01/2020 – 01/2021
- Created efficient data pipelines while adjusting to suboptimal conditions in Central America (rare wi-fi, old tablets, etc.)
- Improved farming outcomes by building a classifier to determine best combinations of pesticides and crops
- Maximized farming data insights by creating Tableau dashboards and teaching leadership/volunteers how to use the dashboards to answer ground operation questions

PROJECTS

Field Experiment on Children's Book "Ennie & Entropy"
- Designed and ran a multifactor field experiment on Facebook Ads to causally infer whether gender representation in online advertisements increases the click-through rate for my self-published STEM children's book "Ennie & Entropy"
- Prefaced the experiment with a power analysis and modeled product performance results with regressions in R Studio

Predicting Scottish Witch Trial Executions
- Researched and preprocessed (SMOTE, PCA) data on historical witch trials and used it to infer whether gender related factors influenced the decision to apply a guilty verdict to the accused witches, which would lead to their execution
- Compared logistic regression to neural net built with Tensorflow and hyperparameters tuned through GridSearchCV

Albanian Topic Modeling Using Word2Vec/BERT/ELMo
- Collected novel tweet dataset using Albanian language research and detected topics in a semi-supervised setting
- Built topic models using embeddings trained on a set of different architectures (Word2Vec/BERT/ELMo) and clustered topics using kMeans. Evaluated results using the tweet hashtags and timestamps, as well as with qualitative analysis

EDUCATION

MS: Master of Information and Data Science. **University of California, Berkeley,** CA   Aug 2022
BA: Data Science. Minor: Business/Industrial Analytics. **University of California, Berkeley,** CA (GPA 3.9)   May 2021
AS: Engineering Science and Mathematics. **Bergen Community College,** Paramus, NJ (GPA 4.0)   May 2019