

EREZ S. SAROUSI

978-235-0335 | ESarousi@yahoo.com | <http://www.sarousi.io> | <https://www.linkedin.com/in/erezsarousi> | <https://github.com/ESarousi>

SUMMARY

A recent masters graduate with an MS in Data Science, I am seeking to utilize my recently gathered data mining, analysis, and machine learning skills to enable data-driven decisions within a dynamic organization and drive growth. I have demonstratable programming and statistical abilities with an analytical aptitude, and collaborative approach.

TECHNICAL SKILLS

Languages: Python (*BeautifulSoup/BS4, JSON, Keras, Matplotlib, NLTK, NumPy, OS, Pandas, Re, Requests, Scikit-learn, SciPy, Seaborn, Sklearn, String, TextBlob*) | R (*caret, datatable, dplyr, foreign, ggplot2, knitr, tidy*) | SQL | HTML

IDEs: Anaconda | Jupyter Labs | Jupyter Notebook | PyCharm | RStudio

Software: Power BI | MS Office (Excel, OneDrive, PowerPoint, Word) | SPSS, Adobe Creative Suite (Acrobat, Photoshop, Premiere) | WordPress

Technical Competencies: A/B Testing | Data Mining & Analysis | Machine Learning | Database Management | Bayesian Statistics | Hypothesis Testing | Logistic & Linear Regression | Random Forest | Decision Tree | Support Vector Machine | K-Means Unsupervised Clustering | Naïve Bayes | Time Series Analysis | Survival Analysis | Data Visualization

LinkedIn Skill Assessment Badges: Machine Learning, Microsoft Excel, R

DATA SCIENCE PROJECT EXPERIENCE

Analyzing Predictors of Stroke

May 2022

- Cleansed and processed stroke and 30+ health parameters using dplyr and caret in R. Conducted exploratory analysis, imputed missing values, and treated outliers. Visualized trends and patterns in the data using ggplot2.
- Implemented five classification models like Decision Tree, Logistic Regression, Random Forest to predict the occurrence of stroke with 83% accuracy and determined four factors that increase risk of strokes.

Virus Anatomy & Computer Defense Algorithm

Apr 2022

- Analyzed ~20k virus and non-malicious files with 15+ features. Explored trends and patterns in the data using Pandas and NumPy. Performed correlation analysis and engineered new features from the data.
- Applied Decision Tree, Random Forest, Logistic regression from Sklearn and SciPy in Python to identify the virus files. Achieved a classification accuracy of 81% and evaluated the efficacy of an ML based antivirus.

Understanding Workplace Turnover With HR Analytics

Mar 2022

- Mined and transformed employee data with 100+ features related to performance, turnover, satisfaction, etc. using Pandas in Python. Generated box plots, line, bar charts, and scatter plots using Matplotlib and Seaborn.
- Performed correlation analysis and linear regression using Sklearn and Statsmodel in Python to uncover potential features related to employee turnover with an RMSE of 12.1.

Data Mining & Text Analysis of Apple Reviews

May 2021

- Preprocessed 10k+ rows of Apple iPhone reviews from Amazon, removed stop words and punctuations, applied lower case formatting, stemmed and tokenized the keywords using NLTK & Textblob in Python.
- Analyzed customer opinions on 10+ phone features like battery, software, and screen quality using sentiment analysis. Identified features that influence customers, showcased insights with Matplotlib visuals.

EDUCATION

Bellevue University | MS in *Data Science*, GPA: 4.0/4.0

Jun 2020 – Jun 2022

Relevant Courses: Statistics for Data Science, Exploratory Data Analysis, Data Preparation, Data Mining, Predictive Analytics, Data Visualization and Presentation, Big Data, Applied Data Science

Worcester State University | BS in *Criminal Justice*

Sep 2010 – Dec 2019

Massachusetts Bay Community College | AS in *Criminal Justice*

Sep 2007 – May 2010

PROFESSIONAL EXPERIENCE

National Grid | *Customer Service Representative*

Apr 2018 – Present

- Investigated root causes for pipeline safety incidents, explored and analyzed relevant data using PowerBI, determined patterns and anomalies in the data using Excel.

Securitas | *Security Officer*

May 2014 - Apr 2018