EREZ S. SAROUSI

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

SUMMARY

I am an accomplished data scientist, with over four years of experience in Python, R, and SQL. My coding-driven automation programs have saved over $100,000 a year by understanding business needs and modeling effective solutions. I am adept at several branches of data science including automation, predictive analytics, data mining, and statistical reasoning.

TECHNICAL SKILLS

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

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

Software: MS Office (including **Power BI**) | **SPSS | Tableau**, **Adobe Creative Suite** | **Salesforce** | **WordPress**

Technical Competencies: A/B Testing | Automated Web Navigation | Bayesian Statistics | Database Management | Data Mining | Data Visualization & Presentation | Hypothesis Testing | Logistic & Linear Regression | Machine Learning | Train-Test Models (Decision Tree | K-Means | Naïve Bayes | Random Forest | Support Vector Machine) | Survival Analysis

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

DATA SCIENCE PROJECT EXPERIENCE

Data Dawg YouTube Channel **Ongoing**

- Hosted a YouTube Channel designed to teach fundamental statistics topics such as central tendency and standard deviation.
- Attracted hundreds of followers and dozens of subscribers by utilizing programs such as Audacity and Adobe Premiere Pro.

Medical Diagnostic Tool **Mar 2023**

- Created a realistic diagnostic tool using Python that simulates patient-facing healthcare scenarios. This program includes human factors to require critical thinking.
- This Python program incorporated packages such as FuzzyWuzzy, Math, Random, and Re
- This tool enables both doctors and those not in medicine to practice their diagnostic skills in a sandbox setting.

Analyzing Predictors of Stroke **May 2022**

- Created predictive models of dozens of health parameters to predict strokes using R packages such as Dplyr and Caret. Processed data, imputed missing values, and treated outliers. Visualized trends and patterns in the data using Ggplot2.
- Implemented five classification models such as Decision Tree, Logistic Regression, and Random Forest to predict the occurrence of stroke with 83% accuracy and determined four factors that increased the risk of strokes.

Virus Anatomy & Computer Defense Algorithm **Apr 2022**

- Analyzed approximately 20,000 viruses 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.
- Applied Decision Tree, Random Forest, and Logistic regression from Sklearn and SciPy in Python to identify the virus files. Achieved a classification accuracy of 81% and evaluated the efficacy of a machine learning-based antivirus.

EDUCATION

Bellevue University | MS in *Data Science* **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**

- Successes: Dean's List – Spring 2018, Fall 2018, Fall 2019

PROFESSIONAL EXPERIENCE

National Grid **Apr 2018 – Present**

- Data Specialist (Distributed Energy Resources, Data Integrity), Process & Support **Sep 2023 – Present**
  - Saved over $100,000 and 2,000 of manual effort annually by developing Python automation tools.
  - Created interactive Power BI dashboards for a company-wide data migration movement that reduced report redundancy by over 80%.
  - Enhanced stakeholder data literacy by creating nearly a dozen video guides for topics such as Salesforce reporting, Power BI, and Large Language Models like Microsoft Copilot and ChatGPT.
- Associate Analyst, Continuous Improvement: Incident Analysis and Corrective Actions **Jan 2022 – Oct 2023**
  - Identified root causes and implemented corrective actions by leading incident analysis meetings.
  - Reduced pipeline safety incidents by over 22% with a trend-identifying and flagging Python program.
- Customer Service, Gas Contact Center **Apr 2018 – Dec 2021**
  - Resolved conflicts to 70+ customers daily, addressing billing concerns to life-threatening emergencies.