Dakota Rossi
Data Analyst

Email : dakotarossi04@gmail.com
Mobile : 515 710 6204

SKILLS

- **Strong**: R, Git, LaTeX, Markdown, Excel, VBA, Regression Analysis, Hypothesis Testing, Statistical Modeling
- **Experienced**: Python, Pandas, NumPy, Scikit-learn, JMP, SQL, Neo4j, Data Structures and Algorithms

INTERNSHIPS

- **Soul Flower** — Summer 2023
  Data Science Intern — *Des Moines, IA*
  - **Application Development**: Created a robust and intuitive CRUD application using Python and Flask, streamlining inventory and product management. This included features for linesheet viewing, file uploads, and PDF downloads, all designed with an emphasis on user-friendly experience. Tailored linesheet outputs to align with customer expectations and ensure satisfaction.
  - **Data Integration**: Harmonized data for various product categories, transforming diverse CSV file formats into a unified schema. Integrated harmonized data into SQLite database, aiding in data retrieval efficiency and manipulation.
  - **Training and Documentation**: Devised a comprehensive user guide for the application and delivered personalized training sessions, enabling non-technical staff to utilize the application effectively. Ensured users felt confident in their understanding of the program and its capabilities.

PROJECTS

- **Football Prediction Model** — 2020-2022
  - **Continuous Development**: Conducted extensive research on various sports databases to find the most complete options. Frequently revisited research, updating dataset selections as model requirements evolved and necessitated additional data.
  - **Data Transformation**: Used R to extract and normalize NFLfastR and cfbfastR package data to generate geographic and performance metrics per team. Used Python for efficient large-scale data transformations.
  - **Model Creation, Analytics**: Developed models using linear regression, random forest regression, a custom KNN regression model, amongst other methods. Tuned models with cross-validation, then combined into classifier for final predictions.
- **Loan Default Classifier** — 2022
  - **Data ETL**: Using Python, transformed data on defaulted loans with Pandas and NumPy for use in modeling and analytics. Data were sourced from MachineHack and Deloitte Global, and focused on bank loans distributed by Deloitte Global.
  - **Modeling**: Used Logistic Regression and KNN classification models on the dataset to predict if loans would default. Tuned models with k-fold cross-verification, performed logistic loss analysis, and created an ensemble model for final classifications.
- **Visualizing COVID-19** — 2023
  - **Data Aquisition and Transformation**: Brought in and transformed Maryland COVID testing data in R for the use of analytics. Data were cleaned for usability in clustering, and adjusted to account for outliers in daily case count.
  - **Visualization and Presentation**: Used R to generate visualizations of COVID-19 case, hospitalization, and death data, normalized and segmented by demographic. Designed visualizations displaying case and death numbers across time, segmented by demographic clusters.
  - **Analytics Summary**: Authored a detailed, in-depth report on methodologies and conclusions, with visuals to help readers better understand the subject. Thoroughly analyzed research limitations and biases related to our findings. Explored implications of findings on the healthcare industry, and potential avenues for further data analysis and insights.

CLUB AND ORGANIZATIONS

- **Sports Analytics ISU**
  Member
  - **Modeling**: Using R and Python for analysis on data from sportsdataverse, created several models based on linear and logistic regression. Models would take sports statistics as input, compare the relative performance of teams, and made predictions of final scores and winning odds across many different sports leagues.
  - **Competitive Analytics**: Used ensemble sports prediction models in inter-organization competitions, where members would make score and standing predictions across various sports leagues and compare results. Continued to expand on sport models, testing limits of regression when compared to other machine learning models.
  - **Club Organization**: Participated in meetings and guest speaker events featuring industry professionals in sports data science. Notably engaged in a detailed discussion and Q+A with the Director of Football Research and Strategy for the Buffalo Bills.
  - **Peer Mentoring**: Provided foundational R programming instruction, collaborated on sportsdataverse API operations, and guided new members on analytics strategies in sports data analysis.

EDUCATION

- **Iowa State University** — Ames, IA
  B.S. Data Science — *2021 - Present*