SPENCER PRENTISS

DATA SCIENTIST

seprentiss@gmail.com 812-572-5678 https://www.linkedin.com/in/spencerprentiss/ https://seprentiss.github.io/portfolio/

SKILLS

Python (Pandas, Numpy, Skit-Learn), R, Java, SQL (MySQL, SQLite), Excel, SAS, Tableau, Hadoop, Spark

EDUCATION

Purdue University **08/2019-12/2022**

B.S. in Data Science and B.S. in Applied Statistics **GPA: 3.68**

Relevant Coursework: Data Mining and Machine Learning, Information Systems, Theoretical Statistics, Probability, Intro to Time Series, Intro to AI, Large Scale Data Analysis, Applied Regression Analysis

WORK EXPERIENCE

Teaching Assistant - Purdue University-West Lafayette, IN 07/2021-12/2022

- Taught Java, Python, GitHub, and Unix environments to 200+ students.
- Collaborated with a team of 20 TAs to address student inquiries, prepare study materials, conduct exam reviews.
- Developed Python exercises teaching fundamental programming, data structures, and web scraping concepts.
- Elevated student comprehension and performance resulting in improved understanding and higher test scores.

Data Analyst Consultant 01/2021-Present

- Utilized Python and Monte Carlo simulations to significantly improve expected win performance (7%) and expected one-possession game performance (13.4%) for a football team looking to beat a division rival.
- Conducted in-depth analysis of election data, developing comprehensive metrics and key performance indicators (KPIs) to guide strategic focus for 3rd party campaign managers.
- Designed and deployed visually captivating data visualizations in Excel, R, and Tableau to facilitate comprehension and data-driven decision-making.
- Delivered multiple engaging presentations to various audiences, effectively offering strategic guidance based on comprehensive data insights across a wide range of contexts and industries.

PERSONAL PROJECTS

Political Science Research Project

- Spearheaded a 6 person research team to investigate the complexity and polarity of judicial opinions.
- Conducted data scraping, cleaning, and analysis of approximately 10GB of opinions from 12 appellate courts.
- Employed advanced natural language processing (NLP) techniques, such as sentiment analysis and LDA topic modeling, to derive meaningful metrics for complexity, polarity, and subjectivity.
- Collaborated with a professor to seamlessly integrate the research findings and methodology into ongoing research initiatives.

Loan Default Analyzer

- Developed a machine learning model using Random Forest Classifier to determine loan default likelihood.
- Conducted feature selection on a dataset with over 1 million data points to identify the most relevant features for model development.
- Employed standardization and sampling techniques, such as normalization and random under sampling, to improve model performance and mitigate bias.
- Successfully built an optimized model with enhanced predictive power, increasing accuracy in determining loan default probabilities.