SPENCER PRENTISS

JUNIOR DATA SCIENTIST

seprentiss@gmail.com 812-572-567 https://www.linkedin.com/in/spencerprentiss/ https://github.com/Seprentiss

SKILLS

Python (Pandas, Numpy, Skit-Learn), Excel, SQL (MySQL, SQLite), Tableau, Natural Language Processing (NLP), R, Java

EDUCATION

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

Bachelor of Science, Double Major in Data Science & Applied Statistics **GPA: 3.65**

Relevant Coursework: Data mining and machine learning, Information systems, Theoretical statistics, Probability, Intro to time series, Intro to AI, Large Scale Data Analysis

WORK EXPERIENCE

Teaching Assistant - Purdue University-West Lafayette, IN **07/2021-Present**

- Introduced over 200 students to Java, Python, GitHub, and Unix environments.
- Developed various exercises in Python to teach the fundamentals of programming, data structures, and web scraping.
- Employed unit testing to grade student assignments.
- Collaborated with 20 other TA's to answer student issues and prepare resources for study sessions and exam reviews.

Data Analyst Intern - Dave Schrader **01/2021-05/2021**

- Directed a partnership with a fellow student in discovering and presenting analytical results, to assist a high school football team in overcoming a division adversary.
- Made use of Monte Carlo simulations in Python to increase expected win performance and expected one-possession game performance by 7% and 13.4% respectively.
- Designed, engineered, and deployed asynchronous and real-time data visualizations in Excel, R, and Tableau.
- Conducted a 45-minute presentation showcasing the findings and communicated how to directly implement them into game strategy.

PERSONAL PROJECTS

Purdue Baseball Analytics **11/2021-Present**

- Coordinated with Purdue Baseball's Assistant Director of Strength and Conditioning to construct dashboards that model athlete performance over time.
- Designed two Tableau dashboards to enable coaches and the 40+ players to see individual and team performance.
- Coaches adapted the weekly training schedules of their athletes based on the data.
- Analyzed 50 games of box score data to discover a correlation between off-field training and on-field performance.

Restaurant Management Web App **02/2022-05/2022**

- Collaborated closely with a team of four other members to create a restaurant management web application.
- Leveraged the Django web app framework to integrate custom-built SQLite database.
- Utilized indexing and Isolation levels to improve performance and consistency.
- Created complex stored procedures and prepared statements to update application data.

Political Science Research Project **01/2022-05/2022**

- Led a six-person team on a research project that investigates the complexity and polarity of Judicial opinions.
- Scraped, cleaned, and evaluated ~ 10GB of data from over 40,000 PDFs containing opinions from the 12 appellate courts.
- Employed NLP techniques such as sentiment analysis and LDA topic modeling to generate metrics for the complexity, polarity, and subjectivity of court opinions.
- Expressed methodology, results, and findings in the form of five 10+ page summaries. The work will be implemented by a professor whose research falls within the field of political science.

Fake Review Classifier **09/2021-11/2021**

- Built a Logistic Regression Classifier that was 97% accurate in predicting whether a review was fraudulent based on user rating, item category, and review text.
- Extracted data on 70,000 lines of text via Natural Language Processing techniques such as Skit-Learn's TDIF Vectorizer.
- Maximized model performance through hyper-parameter optimization and lasso regularization.