Xavier Genelin

Arcadia, WI 54612 | 507-459-0673 | xgenelin@gmail.com | https://xaviergenelin.github.io/

EXPERIENCE

Data Scientist – *Freelance Research*   **July 2022-June 2024**
- Consulted with a medical team to examine rare skin cancer data from SEER registry
- Collaborated using **Github** for version control and sharing code throughout the project Assisted the team in data collection and applying proper statistical and machine learning techniques
- Created a linear regression model in **R** to determine significant variables in skin cancer incidence and write up the results for the final publication

Data Scientist - *Mattress Firm*   **November 2021-July 2022**
- Built customer segments in **python** based on demographic data using scikit-learn to create unsupervised machine learning models to analyze customer habits and look for marketing opportunities for new products
- Utilized **SQL** for data extraction, manipulation, and analysis to support data-driven decision making and machine learning models
- Supported multiple stakeholders across private brand teams in developing data-driven marketing strategies powered by customer analytics and statistical models
- Investigated customer survey data in python with **NLP** using the NLTK library to analyze feedback from various customer groups and determine opportunities for improvement in product delivery and product satisfaction
- Created an **XGBoost** machine learning model to classify customers based on previous transaction habits to aid in customer analysis with **82% accuracy**
- Analyzed the impact of economic stimulus packages on sales, **determining there was an increase**
- Utilized **Github** for version control and collaboration in data science projects while implementing CI/CD pipelines to ensure code quality through automated testing and streamline deployment
- Conducted A/B testing for different media mixed models for new product and private brand teams
- Communicated results with stakeholders (managers to C-suite) to both technical and non-technical audiences

Quantitative Analyst - *NC State Baseball*   **March 2021-June 2022**
- Formed a report using **R** and **Tableau** to analyze NC State pitchers to help optimize their performance
- Consulted the coaching staff based on findings from analysis and assist with game strategy
- Advanced to the 2021 College World Series semifinal and 0.632-win percentage in 2022

Business Intelligence Analyst/Data Analyst - *Ashley Furniture Industries*   **November 2019-November 2021**
- Automated manual processes writing **SQL** queries for data extraction and dashboards, **saving 45 hours per week**
- Designed an app in **python** to optimize the process of diverting shipping containers, **saving 8 hours per week**
- Conducted a statistical analysis in **R** on new product sales and advertisement spending using a machine learning model, determined ad spending had no impact on sales, **saving $300,000**
- Developed and maintained 17 **Power BI** Dashboards for Supply Chain, Manufacturing, Finance, and HR to support data-driven decision making

RESEARCH AND PUBLICATIONS

The Epidemiology of Dermatofibrosarcoma Protuberans Incidence, Metastasis, and Death Among Various Population Groups: A SEER Database Analysis

EDUCATION

Master of Science, Statistics - *North Carolina State University*

Bachelor of Science, Mathematics, concentration in Economics - *Xavier University*

SKILLS

Programming languages: Python, R, PyTorch, PySpark, SQL (T-SQL, MySQL)
Machine Learning/Deep Learning Techniques: Classification, Regression, Clustering, Deep Learning (CNN, RNN), NLP
Tools: Google Cloud Platform (GCP), Jupyter Notebook, RStudio, Google Colab, Tableau, BigQuery, R Shiny

PROJECTS

Terrain Identification [github repo]
- Classify the type of terrain using **PyTorch** from a prosthetic limb based on accelerometer and gyroscope data to create a CNN and tested different hyperparameters for the optimal model

Alzheimer's Risk Factors [github repo]
- Identify risk factors associated with individuals identified with Alzheimer-onset dementia using **PySpark**
- Explore data to determine variables for modeling and use classification models (Decision Tree, Random Forest, K-NN, XGBoost, and Logistic Regression) for prediction

Mobile Game User Retention Testing [github repo]
- Examine user retention for a mobile game using various A/B Tests (a bootstrap, Chi-Squared Test, and a Bayesian Test) in **Python**

NFL Win Prediction [github repo] [shiny app]
- An **R Shiny** dashboard that allows a user to explore NFL game data from 2002-2014 seasons and fit different ML models (Logistic Regression, Classification Trees, and Random Forests) to predict the winners of a game