+1-360-206-7812
Leavenworth, WA
colindmiddleton@gmail.com

Colin Middleton

Entry Level Data Scientist

GitHub: middlec000
LinkedIn: colin-middleton-000
Website: https://middlec000.github.io/

Recent Eastern Washington University graduate with an MS in Applied Mathematics. Completed thesis project on predicting likelihood of homelessness based on an individual's utility billing history. Strong background in applied mathematics and statistics with practical experience in data exploration, manipulation, analysis, and prediction. Seeking to launch career with innovative and learning-focused company.

Skills

Software Languages	Python (pandas, numpy, statsmodels, scikit-learn, matplotlib, seaborn, scipy, tensorflow, streamlit), R, SQL, Java, Git, LaTeX, Markdown
Data Analysis	Correlation, Box plot, Histogram, Q-Q Plot, Hypothesis Testing
Data Preparation	Reformatting, Filtering, Imputation, Class Balancing
Supervised Learning	Linear Regression, Logistic Regression, Cox Proportional Hazards, Vanilla ANN, LSTM, Decision Trees
Unsupervised Learning	K-Means Clustering, Hierarchical Clustering, Principal Component Analysis
Performance Analysis	Performance Metrics, Parameter Interpretation

Education

Master of Science in Applied Mathematics — **September 2019 — June 2021**
Eastern Washington University — *Cheney, WA*

- GPA: 4.0
- First graduate of reopened program (2021)
- Outstanding Graduate Award (2021)
- Graduate Service Appointment (2020 - 2021)

Bachelor of Science in Mathematics — **September 2013 — December 2017**
Western Washington University — *Bellingham, WA*

- GPA: 3.5
- Minors: Chemistry
- Honors Program Graduate
- Math Fellow (2015 - 2017)
- Presidential Scholarship (2013)

Project Experience

Data Science Lead / Homelessness Prediction — **June 2020 — June 2021**
Spokane Predictive Analytics and Master's Thesis Project — *Spokane, WA*

- Established specific research questions.
- Received, matched, and preprocessed de-identified data from multiple files using Python.
- Worked with data providers to iteratively improve datasets.
- Performed data exploration on the 91,591 rows using multiple types of correlation and data visualization to determine relationships within the data with Python.
- Engineered new cumulative features.
- Investigated longitudinal approaches with Linear Regression, Cox Proportional Hazards, and LSTM models as well as non-longitudinal approaches with Logistic Regression and Vanilla ANN models using Python.
- Evaluated and compared models based on their Receiver Operator Characteristic curves. The Logistic Regression model performed best with an Area Under the Curve of 0.81 - similar to current research.
- Utilized K-Folds and minority class oversampling to combat extreme data imbalance (0.39% positive cases).
- Presented findings to the City Council of Spokane and defended as master's thesis.
- Paper pending publication in a peer reviewed journal.

Software Engineer / Academic Paper Clustering by Topic — **January 2021 — March 2021**
Big Data Analytics Class Project, Eastern Washington University — *Spokane, WA*

- Established a data preprocessing pipeline that pulled the body of each paper, detected the language - only retaining English documents, tokenized and lemmatized each text, then converted all documents to sparse vectors of TF-IDF scores using Python.
- Assisted with development of a customized K-Means algorithm optimized for sparse, largely disjoint TF-IDF vectors.
- Developed algorithm component to determine number of clusters to form using the Elbow Method.
- Currently working on publishing a fast hierarchical text document clustering Python package as an extension of this work.

+1-360-206-7812
Leavenworth, WA
colindmiddleton@gmail.com

Colin Middleton

Entry Level Data Scientist

GitHub: middlec000
LinkedIn: colin-middleton-000
Website: https://middlec000.github.io/

Project Experience

Data Analyst / The Wordler **March 2022 — Present**
Personal Project *Leavenworth, WA*

- Acquired data from Kaggle and Wordle website source code.
- Created data filters using regular expressions in Python.
- Created Streamlit website to suggest Wordle words for users: The Wordler

Work Experience

Architectural Drafter **January 2022 — Present**
Alison Miller Architect *Leavenworth, WA*

- Create drawings using AutoCAD LT: floor plans, roof plans, sections, exterior elevations, window and door schedules.
- Final drawing set: dimensioning, code compliance, viewport scaling.
- Site measure existing buildings for remodel.
- Use Dropbox to share files.

Architectural Drafter **October 2018 — December 2021**
Brooks Middleton Architect *Anacortes, WA*

- Created drawings using AutoCAD LT: site plans, floor plans, roof plans, electrical plans, sections, interior elevations, exterior elevations, window, door, and finish schedules.
- Made design adjustments and suggestions.
- Researched building code and construction product specifications.

Related Coursework

Eastern Washington University **September 2019 — June 2021**

- Big Data Analytics (A)
- Advanced Topics in Statistics (A)
- Applied Linear Statistical Modeling (A)
- Data Mining (A)
- Independent Study: Time Series Analysis (A)
- Relational Database Systems (A)
- Data Structures (A)

Western Washington University **September 2013 — December 2017**

- Probability and Statistical Inference (A)
- Nonlinear Optimization (B+)
- Linear Optimization (B-)
- Linear Algebra (A-)
- Mathematical Modeling (A)