

COLLINS WESTNEDGE

✉ Cwestnedge@gmail.com
☎ 281-687-8052
📍 Chicago, IL
in collins-westnedge/
🚲 dirtylittledirtbike

Skills

PROGRAMMING

Python
Javascript
SQL
R
HTML

LIBRARIES

Pandas
Scikit-learn
Numpy
TensorFlow
NLTK
Flask/Dash
spaCy
BeautifulSoup

DATA VISUALIZATION

Plotly
Tableau
Matplotlib
Seaborn
ggplot

MACHINE LEARNING

Natural Language Processing
Regression
Classification
Clustering
Dimensionality Reduction

Education

University of Chicago
Bachelor of Arts Philosophy 2017

School of the Art Institute of Chicago
Continuing Studies Art and Tech 2019

Experience

Metis

Data Scientist

Remote - Chicago, IL
Sept. 2020 to Dec. 2020

Attended a 12-week immersive data science bootcamp focused on project oriented learning. Completed self-designed data science projects from conception to presentation; including data collection, data management, exploratory data analysis, modeling, and visualizations.

Mental Health Online (NLP, Unsupervised)

Analyses Performed:

- Examined how mental health is discussed online as well as the benefits and common issues users face in these support spaces.
- Used non-negative matrix factorization to identify common topics discussed in two of the most active mental health subreddits r/depression and r/anxiety.
- Created vectorized representations of words to examine bias around particular terms prevalent to the mental health community.

Outcomes:

- Provide moderators with information on the daily percentage of posts pertaining to each topic to help them allocate the appropriate resources to provide relevant support.
- Identify and asses posts by urgency, e.g., posts pertaining to the suicide and self harm topics so professional support can be provided in timely manner.
- Insight into community bias around mental health terms so they can make efforts to counteract harmful biases and keep these support spaces inclusive.

Covid Risk Assessment (Web Application)

Analyses Performed:

- Leverages New York Times COVID data for daily number of confirmed cases, as well as BLS data for 2020 population estimates, to assess risks relative to different group/gathering sizes on the county level.
- Establishes risk as the probability that at least one person in a group of a particular size is infected with covid based on the number active COVID cases in a county, an estimation bias term to account for underreporting, and the population size for that county.

Outcomes:

- A live interactive dashboard where users can select any given US county and receive daily estimates of COVID-related risks associated with different group sizes for that location, as well as live reports on the daily change of COVID cases in that county.
- Continues to leverage meaningful insights for organizations and companies to help them actualize data driven decisions about maximum capacity requirements for their offices, store locations, and public venues during COVID.

Toxic Comments (NLP, Classification)

Analyses Performed:

- Examined data gathered from Wikipedia comments as part of a Kaggle competition to identify different forms of toxic speech, such as insults, obscenity, and identity based hate speech.
- Used topic modeling techniques such as non-negative matrix factorization as well as Latent Dirichlet Allocation to engineer additional features for classifiers.
- Trained naive Bayesian classifiers to identify and predict different forms of toxic speech and examined model coefficients to identify bias towards LGBTQ terms in the training data.

Outcomes:

- Created classifiers to identify and flag different forms of hate speech in online forums with an f1-score of .65 and 95% accuracy on unseen data.
- Provided insights on how legitimate and non-offense terms specific to marginalized communities occur frequently in the context of hate speech.
- Developed strategies to counteract these biases in the data to prevent models from flagging or censoring legitimate posts by members of these communities.

Predicting Housing Prices (Regression, Web Scraping)

Analyses Performed:

- Collected information on properties from Zillow pertaining to the greater Chicago area by leveraging BeautifulSoup to collect and parse data based on specific search queries.
- Analyzed 22 unique property features for 2600 properties.
- Engineered additional geospatial features to improve model performance.

Outcomes:

- Reusable python script for efficiently gathering large amounts of housing data from the Zillow page that can easily work around numerous web page limitations.
- Multiple models which help determine the fair market value of a home.
- Best performing model correctly estimates the fair market value of a home within a \pm 35,000 dollar margin.

Hanover Investment Advisors

Statistical Modeling and Programming Analyst

Chicago, IL
Sept. 2015 to Oct. 2017

Worked with commercial real estate data and provided statistical analysis and insights in R for presentations to U.S. and European pension funds, pension fund advisors, insurance companies, sovereign wealth funds as well as U.S and European banks and Singapore family office.