

ALLEN NI

✉ azni615@gmail.com
☎ 847-924-5350
📍 Vernon Hills, IL 60061
in allen-ni-7396a9193/
🌐 azni615

A data professional who aims to bring multiple fields together to create novel solutions and compelling stories that improve the world through engineering, and analysis. Creatively bringing a multidisciplinary outlook to complex problems to find simple, yet nuanced solutions that improve the lives of those involved.

Skills

DESIGN

Six Sigma Green Belt
Iterative Design
Systems Modeling
Rapid Prototyping
Project Management
A/B testing

DATA SCIENCE LIBRARIES

Pyspark
Pandas
Numpy
Scikit learn
Keras
Statsmodels
Wordcloud
Git

SOFTWARE

Fusion 360
Excel
SQL
Python
MATLAB
Java

DATA VISUALIZATION

Seaborn
Matplotlib
Plotly
Tableau

MACHINE LEARNING TECHNIQUES

TF-IDF
Word Embeddings (Word2Vec)
NLP
LDA
NMF
CountVectorizer
Kmeans clustering
SVD
PCA
Neural Networks

CLOUD/DATA STORAGE

AWS
SQL
Postgres
MongoDB
Hadoop

Experience

Metis

Data Scientist

Chicago, IL
June 2020 to Sept. 2020

- Metis is an ACCTE accredited 12week immersive data science bootcamp focused on project oriented learning.
- The core curriculum is centered around Python, statistics, supervised and unsupervised machine learning, exploratory data analysis, databases, and visualization techniques.
- Completed five self-designed data science projects from conception to presentation; including data collection, data management, exploratory data analysis, modeling, and visualizations.
- Select project highlights are in the projects section.

FoodVaccine Inc

Research/Product Assistant, Adviser

Urbana, IL
Oct. 2018 to Dec. 2019

Consulted on and helped develop research regimens and product targets with Foodvaccine. Facilitated interpersonal and inter-organizational networking, market research, technical research, and grant applications.

Cardinal Bernadin Cancer Center, Loyola University Medical Center

Head Intern/Research Assistant

Maywood, IL
June 2015 to Sept. 2016

Lead and trained 3 other interns while managing lab resources, directing research, and conducting lab work under Dr. John Xin. Published 3 papers with the lab.

Research Intern

Conducted experiments, lab work, proofreading, and presentations under guidance from PIs and fellowship students in a BSL-2 lab.

Maywood, IL
Mar. 2014 to May 2014

QLPOINT INC.

Assistant Analyst

Vernon Hills, IL
Sept. 2013 to Mar. 2014

Conducted secretarial and accounting work for Financial/consulting firm QLPOINT, later moved into assisting with the financial analysis of target markets and companies.

Mathnasium Vernon Hills

Head Tutor

Vernon Hills, IL
June 2012 to Sept. 2013

Head tutor for math students K-12, created systems and protocols to set up business operations, customer relations, and tutor training.

Projects

MTA Charity Recruitment

June 2020 to July 2020

- Optimized recruitment efforts based on NYC MTA turnstile data.
- Obtained 2 years of MTA turnstile data from NYC MTA data portal. (2018 to 2020).
- Processed in Pandas then mapped with Folium.
- Compared Pre and Post COVID data to recognize underlying human patterns that were not based on traditional commuting. Matched high-throughput stations with COVID cases and found correlations.

Web Fiction Content Visibility

July 2020

- Used NLP and Web Scraping to determine correlations between different metrics of Web Fiction and their success.
- Data was obtained by webscraping the Harry Potter section of Fanfiction.net (over 20000 individual works analyzed)
- Created a logistic regression with LASSO in SciKitlearn and used NLP (vectorizers) for more features.
- Results indicated that "Quality" of web-fiction is less important than regular update schedules and number of updates per length of time. Works labeled for mature audiences also get more engagement and are therefore more successful for less effort.

Law School Classifier

July 2020 to Aug. 2020

- Used Demographic and Historical pass rate data from the American Bar Association to predict if Bar Exam pass rates will be above a threshold.
- Created a SQL database that held American Bar Association school disclosure data from 2009 to 2019.
- Used SQL to stitch together data from 8 different disclosure tables. Created two models, one based on gradient boosted trees, another a tuned logistic regression. Used Tableau to visualize pass rates and other metrics on a per-state basis.
- Both models indicated that pre-Law school academics and results were the most important factor to determine Bar Exam average (better schools get better students, who inherently do better on the exam). Number of part time students has a strong negative correlation. Admissions ratio, International student ratio, and financial aid received all had no strong correlations.

NSF Grant Abstract Topic Extraction

Aug. 2020

- Used NLP to extract topics for 3 different classes of approved NSF grants to find patterns.
- Ran analysis on 35k grants from a dataset of approved NSF grants from 1979 to 2016.
- Used vectorizers and TF-IDF to extract topics and wordclouds to visualize differences between the 3 classes of grants.
- Results indicated that most approved grants contained many contingencies and smaller goals. Further analysis indicates that the NSF tended to approve grants that would have results whether the project succeeded or not, and avoided heavy investment for necessary equipment that labs didn't have already.

Cross Media Recommender

Sept. 2020

- Used NLP to create a recommender that matches Video queries with Book recommendations and vice versa.
- Web Scraped to obtain Video media data from theMovieDB.org and Netflix/IMDB, and book data from BookCrawler utility, Google Books API, and Goodreads API.
- Used Word Embeddings (with Word2Vec), TF-IDF, and jaccard distance to create a content recommender.

Education

University of Illinois, Urbana-Champaign
BS Bioengineering 2019

Therapeutics/Nanotechnology focus, with experience in design, medicine, vulnerable populations, and international entrepreneurship & outreach (Africa)