

# ANKITA PATIL

Data Scientist

patilankita2491@gmail.com  
8155086376  
Denver, CO



## EDUCATION

Certification  
Data Scientist  
**General Assembly**  
January 2022 – Current  
Remote

M.S.  
Electrical Engineering – Power Electronics  
**Northern Illinois University**  
August 2013 – August 2015  
Dekalb, IL

## SKILLS

**Programming & Analytical Tools**  
Python (NumPy, Pandas, Scikit-learn, Sci-Py, Matplotlib, Seaborn) TensorFlow  
SQL (MySQL, Postgres)  
Tableau  
GitHub  
Web Scraping  
Streamlit  
Excel  
Matlab

**Machine Learning**  
Linear Regression  
Classification  
K-Nearest Neighbors  
K-Means Clustering  
DBSCAN  
Decision Trees  
Ensemble Learning  
Neural Networks  
Predictive Modeling  
Sampling Methods

## SUMMARY

Analytical and detail-oriented Data Scientist with 5 years of experience as Electrical Engineer. Looking to combine my passion for bringing designs to life and solving problems by helping businesses drive strategic decisions. Proven ability to understand business requirements and translate complex needs into viable solutions.

## PROFESSIONAL EXPERIENCE

Data Science Fellow

**General Assembly**

January 2022 – April 2022

Remote

- Full-time fellow in Data Science Immersive program with focus on real-world application of principles and best practices in data science in a 500-hour immersive course
- Developed skills in data gathering and wrangling, data analysis, supervised and unsupervised model techniques with structured and unstructured data, and data visualization over the course of individual and collaborative projects
- Determined appropriate strategies to improve data quality, impute missing data, feature selection and optimize & evaluate model performance

Electrical Design Engineer

**Olsun Electric Corp.**

September 2015 – March 2020

Richmond, IL

- Designed custom-built transformers for utility, PV systems, EV charging stations and industrial systems
- Troubleshoot performance issues using computer-based models and empirical data
- Collaborated with Sales, Engineering, Manufacturing and Testing to transform customer inputs into robust designs for production

## PROJECTS

Hybrid Movie Recommender

- Aggregated data from MovieLens, IMDb and TMDB to build a hybrid movie recommender by ensembling content-based and collaborative filtering
- Implemented content-based filtering with movie features using NLP and Sentiment Analysis and user-based collaborative filtering
- Designed a streamlit app using the recommender where users rate the movies they watched and get personalized movie recommendations

Classification of Unstructured Data Using Logistic Regression

- Scraped posts from Reddit's Pushshift API to build a model that identifies whether a question is more likely to come from r/explainlikeimfive or r/AskEngineers based on their semantic content
- Created Logistic Regression and Random Forest classification model using NLP with 53% improved accuracy over the baseline model and achieved 92% precision and 93% recall

Clusters of Counties in COVID-19 Using K-Means Clustering

- Investigated disparity in COVID-19 vaccination across the United States by clustering counties based on vaccine responses, demographics and socioeconomic data and deriving insights into the ideological divide in the country utilizing K-Means Clustering
- Aggregated this data to predict 2020 election results using Ensemble Model (KNN, Random Forest, Bagging and Ridge) where model exhibited 90.3% accuracy

House Price Prediction Model Using Linear Regression

- Gathered housing data from Ames, IA to determine important features that sell a house and developed algorithm that will predict the price of a house in Ames, IA
- Built Linear, Lasso and Ridge Regression models reviewing R2 and RMSE score, with the best model improving RMSE score over baseline model by 28%