

Thomas Zou

**626 608 8363**

**thomas.zou97@gmail.com · linkedin.com/in/thomas-zou-bb057b229/**

# Summary

* **Certified** in **Data Science, Machine Learning, and Deep Learning** from **IBM**.
* **Around 2 years of experience** as a **Machine Learning Engineer** and **Data Scientist/Analyst.**
* Hands-on experience in data design and analysis using **Machine Learning Techniques** and modules.
* Experience working with **MySQL** and **Oracle 18C** databases
* Strong knowledge in writing **SQL Queries, sub-queries, and joins.**
* Strong background in **Machine Learning, Predictive Analysis and Data Mining** with a broad understanding of **Supervised and Unsupervised learning** techniques and algorithms (e.g.: **Linear and Logistic Regression, K-NN, SVM, Naive Bayes, Decision trees, Clustering, K-Means, Mean Shift, Hierarchal Clustering, DBSCAN, Decision Trees, Random Forest**, etc.)
* Experience working with **Artificial Neural Network (ANN)** and **Convolutional Neural Network (CNN)** and **Recurrent Neural Networks (RNN)**
* Strong knowledge in **Time Series Analysis**, such as **ARIMA** and **SARIMA** model implementation.

# CertificatIONS

* **AZ-900 —** Microsoft Azure Fundamental
* **IBM DSN —** Data Science Tools
* **IBM DSN —** Deep Learning using TensorFlow
* **IBM DSN —** Deep Learning Essentials
* **IBM DSN —** Data Visualization using Python
* **IBM DSN —** Data Science Methodologies
* **IBM —** Machine Learning with Python
* **IBM —** Python for Data Science

# Skills

**Basic:** Algebra, Calculus, Statistics Data Structures, Data Science, Algorithms, Machine Learning, Deep Learning, Azure, and Power BI

**Languages**: Python, JavaScript, SQL, MongoDB, HTML, Django, Tkinter, and CSS, C++

**Libraries:** NumPy, Pandas, Matplotlib, Seaborn, Sklearn, XGBoost, LightGBM, CatBoost, Pylab, TensorFlow, Keras, NLTK, word cloud, and Stats Models

**IDEs/Development Tools:** Jupyter Notebook, Colaboratory, Spyder

**Machine Learning Algorithms:** Linear and Logistic Regression, Decision Tree, SVM, Naïve Bayes, KNN, K-Means, Mean Shift, Hierarchal Clustering, DBSCAN, Time Series, Random Forest, Voting Classifier, AdaBoost, XGBoost, CatBoost, and LightGBM

# Experience

|  |
| --- |
| June 2021 – PresentData Scientist/Machine Learning, **Synergistic It** I have been responsible for applying the CRISP-DM process on datasets and business projects. I excel in data exploration and modeling, data analysis, data manipulation, preforming machine learning and deep learning algorithms, optimizing algorithms, creating project reports, and creating graphical dashboards. |

# Projects

**Product Sales Prediction (July 2022 – Present) Regression**

**Data Scientist/Machine Learning Engineer**

**Project Description**: Predict the purchase amount for various products bought from a specific user and their provided user information. The dataset consists of categorical and discrete values for the user’s age range, gender, occupation, residence, and product category. Apply multiple linear regression to find collinearity amongst the features and predict the continuous outputs. Preform Ridge and Lasso regression to predict a generalized model.

**Tools Used:**Python, Power Bi, Flask

**Libraries Used:**Pandas, NumPy, Matplotlib, seaborn, Scipy, Sklearn, Pylab, stats models, LightGBM

**Algorithm used**: Multiple Linear Regression, LightGBM Regressor

**Roles and Responsibilities:** I was responsible for data preprocessing (data cleansing, Feature Engineering), feature encoding, and setting up the data for model implementation. Build Power Bi Dashboard to visualize and preform EDA on the data. Created MLR model and checked for collinearity between all features.

**Company Reviews Analysis (Feb. 2022 – June 2022) NLP and Boosting**

**Data Scientist/Machine Learning Engineer**

**Project Description**: A set of reviews that has been voted as offensive or constructive criticism. Most of the bad reviews consisted of offensive language which have been flagged and submitted for more scrutiny. My team and I must identify which of these comments are correlated with constructive criticism or just offensive speech to assess whether future reviewers should be flagged or heard by the company. We will accomplish this using sentiment analysis and various classification models.

**Tools Used:**Python, html

**Libraries Used:**Pandas, NumPy, Matplotlib, seaborn, re, Sklearn, word cloud, XGBoost, LightGBM

**Algorithm used**: SMOTE, Random forest, Decision tree, Logistic regression, KNN, Voting classifier, SVC, Naïve Bayes, XGBoost, LightGBM

**Roles and Responsibilities:** I was responsible for data preprocessing (data cleansing and Feature Engineering) and setting up the data for model implementation. I also created word clouds for constructive criticism, offensive, and both.

**Dog Breed Classification (Nov. 2021 – Jan. 2022) CNN**

**Deep Learning Engineer**

**Project Description**: Predict the breed of dogs from their image. The dataset consisted of image ids and a separate file contained all the various images of dogs labeled by their image id. The data set should be trained using convolutional neural network algorithm. Multi-classification is required since there are multiple dog breeds to identify from.

**Tools Used:**Python

**Libraries Used:**Pandas, NumPy, TensorFlow

**Algorithm used**: Convolutional Neural Network, Artificial Neural Network

**Roles and Responsibilities:** I was responsible for importing the dataset and the file of images so that each image matched a corresponding image id in the dataset. I split the data set into train and validation (test). I applied a convolutional neural network on the training images. I also reduced the sizes and dimensions of the images using pooling and flattening layers. Lastly, I applied an artificial neural network onto the flattened image to predict the result of the breed.

**Credit Card Fraud Detection (Sep. 2021 – Nov. 2021) Classification**

**Data Scientist/Machine Learning Engineer**

**Project Description**: Predict whether a user’s credit card account is classified as fraudulent or not. The dataset contains many features. The dataset must be tested using various classification algorithms to find the best result and understand why some algorithms are better than others. Create a confusion matrix and graph for each algorithm used. Preform oversampling and under sampling to improve the model.

**Tools Used:**Python

**Libraries Used:**Pandas, NumPy, Matplotlib, seaborn, Scipy, Sklearn, collections, stats models

**Algorithm used**: Principal Component Analysis, SMOTE, Under sampling, Naïve Bayes, Logistic Regression, KNN, SVM

**Roles and Responsibilities:** I was responsible for applying the various algorithms listed on the model without sampling techniques. I preformed PCA and found the most important features to display most of the data while lowering the size of the dataset. I built the template for my team to display the confusion matrix and classification tests for the various algorithms used.

**Bakery Sales Analysis (June 2021 – Sep. 2021) Association Rule**

**Data Scientist/Machine Learning Engineer**

**Project Description**: Identify bakery items that were frequently purchased together using market basket analysis. Compute lift for items to find the highest association between frequently purchased items using association rules. Analysis data to find the most profitable hours of operation. Use association rule to find the most frequent antecedent and consequent pairs. Those pairs with high confidence and conviction scores, are that which would more likely occur if the antecedent was picked first.

**Tools Used:**Python

**Libraries Used:**Pandas, NumPy, Matplotlib, seaborn, MLX tend

**Algorithms used**: Market Basket Analysis, Association Rule, and K modes clustering

**Roles and Responsibilities:** I was responsible for end-to-end implementation viz. primary data collection, data preprocessing (data cleansing, Feature Engineering), and setting up the data for the various models used.

# Education

### Graduated: December 2021

## Engineering Physics-Minor in Computer science,

## University of Colorado, Boulder

**Bachelor of Science**