

**Mayura Sunil Zadane**

**Data Scientist - AI**

3+ years experienced Python Developent and Data Scientist with the ability to apply ML techniques and algorithm development to solve real-world business problems. Highly adept at machine learning technique clustering and classification, statistical modeling, NLP, text analysis, data analysis and data visualization to increase business efficiency.

* A detailed oriented professional with 3+ years of experience in Python, Machine Learning and Data Science with expertise in Financial Domain.
* Ability to achieve in-depth understanding of the problem domain and available data assets.
* Able to investigate Data Visualization and summarization techniques conveying key findings.
* Communicates findings and obstacles to team members to achieve best approach.
* Experience in Web Framework Flask.
* Experience in data management tools - Relational and SQL databases, MongoDB.
* Knowledge of Python’s Data Analysis and Machine Learning Libraries.
* Ability to use Web Scraping Tools as Beautiful Soup.
* Data mining algorithm experience in the families of predictive algorithms ( Regression, KNN, Decision Trees) and clustering algorithms (k-means and Hierarchical Clustering). Implementing regularization techniques like Lasso and Ridge in regression. Working on Time Series Data and performing Time Series Analysis.
* Ability to process Text Processing using NLTK library and other Natural Language Processing techniques.
* Implementation of Recommendation system using Collaborative and Content based filtering.
* Strong communication and interpersonal skills.

# **Skill Set**

**Machine Learning & Data Science:**

* Experience of Machine Learning and Deep Learning Fundamentals.
* **Python/ML/CV Packages:** Scikit Learn, Pandas,Numpy, Matplotlib, Seaborn, Scipy
* **Text Processing:** NLTK, Term Frequency-Inverse Document Frequency (TF-IDF), Bag of Words
* **ML Classification :** Linear Regression, Logistic Regression, Lasso & Ridge Regression, Naive Bayes Classifier, k-NN, Support Vector Machines, Decision Trees, Random Forest., Gradient Boosting ,AdaBoost, XGBOOST and Clustering

**Supporting Technologies:**

* **Programming Languages :** Python, R, C++ ,SQL
* **Databases :** SQL, NoSQL
* **Web stack :** Flask
* **Cloud Platforms :** AWS
* **DevOps :** Git
* **Operating System :** Windows, Linux

# **Recent Projects**

**Project 1: Loan Approval and Defaulter Prediction**

**Domain:** Finance

With the enhancement in the banking sector lots of people are applying for bank loans but the bank has its limited assets which it has to grant to limited people only. Banks gives money to people in exchange for the promise of repayment. Some will default on the loans, being unable to repay them for some reason. The main objective was to predict which people will default on their loans based on their financial information.

**Project 2: Document Classification**

**Domain:** Natural Language Processing

Document Classification can help an organization to meet legal and regulatory requirements for retrieving specific information in a set timeframe. Document classification is an example of Machine Learning (ML) in the form of Natural Language Processing (NLP). By classifying text, class can be assigned to each page of the document. It involves Text-Preprocessing, Dynamic Language modeling, Natural Language Processing.TF-IDF is used to find the features of intra-class documents. A statistical model is used to find and differentiate the inter-class features. K-Nearest Neighbor, Support Vector Machine, Naïve Bayes classifiers are used to classify the documents.

**Data Science Intern**

* Gained a solid understanding of machine learning and data mining concepts with supervised and unsupervised learning methods such as Regression, Support Vector Machine, Naive Bayes, Decision Tree, PCA, Clustering, Apriori Algorithmetcs
* Also learned advance machine learning Algorithms like Random Forest, Adaboost, Gradient Boost, XGBoost.
* Learned to design lean proofs of concepts to answer targeted business questions. Explored and worked with a wide range of proprietary, interesting data stores performed ETL & learned to apply existing algorithms
* Completed general practice on Iris data-set, Titanic data-set, Breast cancer, PIMA Diabetes, EDA.

**Responsibilities**

* Mine and analyze data from company databases to drive optimization and improvement of product development, marketing techniques and business strategies.
* Work with stakeholders through the organization to identify opportunities for leveraging company data to drive business solutions.
* Search for ways to get new data sources and access their accuracy.
* Access the effectiveness of new data sources and data gathering techniques.
* Use predictive models to increase and optimize customer experiences, revenue generations and other business outcomes.
* Coordinate with functional teams to implement models and monitor outcomes.
* Develop processes and tools to monitor and analyze model performance and accuracy.