

ADITH GUSAIN

Mobile: 9368035192 / 7017834178

Email: adith.gusain999@gmail.com

CAREER OBJECTIVE

A passionate **Data Scientist** with 3+ years of work experience in delivering client projects using techniques like **Regression/Classification, supervised and unsupervised algorithms**, and **NLP**. Hence, I am seeking a challenging position in the area of **Data Science, Machine Learning, Data Analytics** and **Artificial Intelligence** where I can utilize my skills to deliver insights and implement action oriented solutions to complex business problems.

WORK EXPERIENCE

Deloitte, Pune

AI Developer

September, 2019 – Present

Roles and Responsibilities–

- Analyzing client demand, conducting exploratory data analysis to extract insights and uncover trends and relationships in data, and also performing **Data Visualization**, **Data cleaning**, and **Building Machine learning models**.
- Data Cleaning & Data Preparation: **Missing value imputation**, **outlier treatment**, and **Exploratory Data Analysis (EDA)** over continuous & categorical variables.
- Work with a senior developer to manage large, complex data projects for corporate clients.
- Complete detailed programming and development tasks for the back-end, as well as challenging back-end server code.
- Data manipulation, Data wrangling, Statistical analysis**, and **Feature selection** in order to decrease the RMSE and get an optimized model.

PROFESSIONAL SUMMARY

- Expertise in **Python & Data Analytics, ML modeling, NLP**, and a proven ability to manage complex tasks.
- Proficient working knowledge in managing the entire Data Science project life cycle and actively involved in all phases, including **Data collection, Data Preprocessing, EDA, statistical methods, model building**, and **model validation**.
- Experience with data management tools like **MySQL** and **MongoDB**.
- Experience in Python web application development using the **Flask Web Framework**.
- Exposure to key concepts in NLP like **Stemming, lemmatization, Bag of words, Text Classification, Topic Modeling, TF-IDF, Word2Vec**, and **Cosine Similarity**.
- Expertise in manipulating and analysing complex, high-volume, high-dimensional data from various data sources
- Experience deploying the project on a cloud platform like **AWS** by creating EC2 instances.
- Have a good understanding of concepts in **Deep Learning** like **Optimization, Activation, ANN, CNN, RNN**, etc.
- Solid understanding of the mathematics related to data science like **probability, statistics, linear algebra**, etc.
- Ability to write clean and production-ready code with **Object Oriented Programing** in Python.
- Soft Skills:** Good Communication, Strong Decision Maker, Complex Problem Solver, Creative Thinker, Innovative, Service-focused.

TECHNICAL SKILLS

Languages: Python, SQL

Libraries: Pandas, Numpy, Seaborn, Tensorflow, Keras, Matplotlib, Regex, Sklearn, NLTK, Spacy, etc

ML Algorithms: Linear Regression, Logistic Regression, Lasso and Ridge Regularization, SVM, Naive Bayes, Decision Tree, Random Forest, K-means Clustering, Hierarchical clustering, K-Nearest Neighbour(KNN), PCA, AdaBoost

Operating Systems: Windows, Linux

Tools & Repository: Anaconda, Jupyter Notebook, Visual Studio, Git

EDUCATION

- **Bachelor of Engineering** with **73.6%** from **Graphic Era Hill University**, Dehradun, Uttarakhand in **2019**
 - **Higher Secondary** with **80.8%** from **Army Public School**, Dehradun(CBSE) in **2015**
 - **Secondary School** with **77.9%** from **Army Public School**, Dehradun(CBSE) in **2013**
-

PROJECTS

PROJECT 1

Sentiment Analysis and Multiclass Classification of Customer Reviews:

Sentiment Analysis of reviews on an online shopping site where clothes of various brands are available. And find the root causes and categorising them into different classes and building an application to identify the issue and send notifications to respective departments.

Responsibilities:

- To perform sentiment analysis on the reviews as well as automate the manual process of classifying reviews.
- Employed essential **text preprocessing** techniques such as **text cleaning**, **tokenization**, **stop word removal**, and **lemmatization** using **Spacy** and **NLTK**
- Performed **Feature extraction** by using the Word2Vec pertained model.
- Find the root causes and map them to their respective categories using **cosine similarity**
- Build a **Multiclass Classification** model to classify the negative reviews into quality issues, delivery issues, etc. by using algorithms, namely **Random Forest**, **SVM**, **Naïve Bayes**

PROJECT 2

Optimal Price Prediction for the Homestays:

Our client was a broker company. They have a website where people can book hotels and homestays at an optimal price. We need to predict the optimal price for the stay, get some insights about the provider's performance, and understand the customer's needs.

Responsibilities:

- Collected business requirements by communicating and coordinating with the respective teams and individuals.
- Data Wrangling, analysis, and discovering patterns using **Pandas**, **NumPy**, **Seaborn**, and **Matplotlib** and gave insightful recommendations to business leaders.
- Features engineering such as **Feature Scaling**, **Outlier detection**, and **Data visualization** through **Standard Scaler**, **bar plot**, **box plot**, and **histogram**.
- Performed **Hyperparameter tuning** to optimize accuracy and loss.
- Created an API using the **Flask Web Framework**.

PROJECT 3

Document Classification:

Classification of various medical documents from a private hospital. It involves text extraction from the image, natural language processing, and TF-IDF is used to find intra-class features. Support Vector Machine and Naïve Bayes classifiers are used to classify the documents.

Responsibilities:

- Developed a Tesseract OCR module to extract text from the images.
 - **Text preprocessing** and cleaning using the **NLTK** library.
 - Generated sparse matrix using **TF-IDF Vectorizer** by tuning parameters.
 - Build a classification model using **SVM**, **Naive Bayes**, **Random Forest**, and **Logistic Regression**.
 - Validated and selected models using **k-fold cross validation**, a **confusion matrix** and worked on optimizing models for high accuracy.
-

PERSONAL DETAILS

DOB: 09/01/1997

Address: Ajabpur Khurd, Banjarawala Road, Dehradun, Uttarakhand 248001

Languages: English & Hindi

Marital Status: Unmarried

Hobbies: Playing Guitar, Travelling, Trekking