

Aarushi Gupta

Data Science Intern

PERSONAL INFORMATION

Phone Number
9958276702

Email Address
aarushi.gupta17@gmail.com

LinkedIn
<https://www.linkedin.com/in/aarushi-gupta-63494b21b/>

Github
<https://github.com/Aarushi17Gupta>

EDUCATION

International Institute of Information Technology Bangalore
Post Graduate Diploma
Machine Learning and Artificial Intelligence
(December 2021 - Present)

Kumaun University, Nainital
Master of Science
Physics
(July 2018 - November 2020)
Percentage- 80.2%

Amity University, Noida
Bachelor of Science Honours
Physics
(July 2014 - June 2017)
CGPA- 9.71

AWARDS

Gold Medal
Kuamun University, Nainital

First in the university for the academic performance in M.Sc. (Physics)

Gold Medal
Amity University, Noida

First in academic performance in B.Sc. (Hons.) Physics

CERTIFICATIONS

The Data Science Course 2021: Complete Data Science Bootcamp
Udemy

The Complete SQL Bootcamp 2022: Go from Zero to Hero
Udemy

Machine Learning Certification Course for Beginners Analytics Vidhya
Analytics Vidhya

Fundamentals of Data Visualization in Tableau
Udemy

Python Basics
University of Michigan | Coursera

What is Data Science?
IBM | Coursera

SKILLS

Languages : Python	Database: PostgreSQL, MySQL
ML Algorithms: Linear/Logistic Regression, Clustering, Decision Tree	NLP: Text Normalization (Stemming and Lemmatization), Sentiment Analysis, Word Cloud
Data Visualization: Matplotlib, Seaborn, Tableau	Libraries: Pandas, Numpy, Scikit-learn, tensorflow (basics)

SUMMARY

I am an aspiring data scientist who is in search of a dynamic workplace where I can constantly learn new techniques and grow professionally. My educational degree has prepared me to take on mathematical and statistical problems and solve them as efficiently as possible. Sincerity, commitment and analytical problem solving skills are my forte which I believe will prove beneficial to any organization that I will be a part of.

WORK EXPERIENCE

Alpha AI
Data Science Intern
(November 2021 - Present)

- Working on **LawDiktat**: AI enabled project to help lawyers and other legal teams to recommend similar legal documents using various algorithms.
- Working on various keyword extraction and summarization techniques for document recommendation system such as keyBert, TextRank, TopicRank, Rake, Yake and Scapy.
- Working on **Tripsero** :Performing EDA, Data Cleaning and Data Merging on various datasets. Also, assigning tags to each data so as to create a recommendation engine which is personalized to each user.
- Building a recommendation system by using clustering and knowledge graphs for recommending trip destinations to the users.

Technocolabs Softwares
Data Analyst Intern
(January 2022 - Present)

- Using Exploratory Data Analysis on Mortagage Backed Securities on 'Freddie Mac's Single Family Loan-Level Dataset.
- Predicting **Mortgage Prepayments** in the industries by empirical research and statistical Modelling.
- Using a number of variables to predict prepayment risks such as borrowers' credit score, income, loan-to-value ratio, income, and loan age.

The Machine Learning Company
Machine Learning Fellow
(November 2021 - Present)

- Creating the app for the machine learning model and using Streamlit in the local system. Also deploying the app through **Streamlit Sharing**.
- Building a General Purpose Recommendation system using cosine similarity to choose similar users and similar items to the existing user.
- Using a User based, Item based and Model based collaborative filtering using K nearest neighbors model and designing a hybrid model for both the content based and collaborative based models.

The Sparks Foundation
Data Science and Business Analytics Intern
(November 2021 - December 2021)

- Created algorithms based on deep-dive statistical analysis and predictive data modelling.
- Used unsupervised machine learning and created a clustering model for IRIS dataset with K-means clustering algorithm.
- Implemented a decision tree classification model to classify the IRIS Dataset.
- Performed Exploratory Data Analysis on 'Global Terrorism' dataset to draw insights and find out the hot zones of terrorism.

LetsGrowMore
Data Science Intern
(November 2021 - December 2021)

- Used programming to convert an image into pencil sketch using cv2 library of python.
- Deployed quantitative analysis and data visualization to present data. Extracted, processed and analyzed large datasets to solve analytical issues.
- Developed a deep neural network machine learning project to read handwritng for MNIST handwritten digit classification challenge using Tensorflow.

Aryabhatta Research Institute of Applied Sciences(ARIES)
Research Intern
(May 2016 - June 2016)

- Used Python programming language for analysing the data for various events happening in space.
- Worked on 'Time Series CCD Photometry of Exoplanet'.

PROJECTS

General Purpose Recommender System Deployment using Streamlit
<https://share.streamlit.io/aarushi17gupta/recommender-system-deployment-using-streamlit/main/recommend.py>

Created a General purpose Recommender System that can recommend any items such as Movies, Receipes, Books and MOOCs. Used cosine similarity, TF-IDF Vectorization and K Nearest Neighbours to recommend top items from a list of large items. At the end, deployed the machine learning model using Streamlit Sharing to create a Recommender System application.

Twitter Sentiment Analysis

Used sentiment analysis to analyze the tweets and segregate them into various segments like positive, negative and neutral tweets with a model accuracy of 82%.

Sentiment Analysis on Yelp Reviews

Used EDA Augmentation and various algorithms such as Random Forest classifier, Logistic Regression, XGBoost, Naive Bayes and Support Vector Machine (SVD) for comparing accuracies and other evaluation metrics.

Document Similarity Measures

Measured similarity between text documents by treating the text the words in the documents as vectors and using various similarity measures such as FuzzyWuzzy, Cosine Similarity and Jaccard Similarity.