Ankit Gupta

(+91) 8130102072
ankitgupta1729@gmail.com
https://github.com/ankitgupta1729
https://gateoverflow.in/user/ankitgupta.1729

Education

2019 - 2021	M.Tech(CS), *Indian Statistical Institute, Kolkata*, 80.90 %
2011 - 2015	B.Tech(IT), Ajay Kumar Garg Engineering College, Ghaziabad, 71.90 %
2010	Intermediate (U.P. Board), Shivaji Inter College, Kanpur, 82.60 %
2008	High School (U.P. Board), Shivaji Inter College, Kanpur, 77.00 %

Work Experience

- 3-years experience (Since 2^{nd} August 2021) with Reliance Industries Limited (RIL) as a Data Scientist.
- 3-years experience (from Nov, 2015 to Dec, 2018) as SQL developer and database administrator in Wipro,Bangalore.
- Internship from Hewlett Packard (HP) in 2014 for 45 days.

Master Thesis

Problem: Boolean Function Approximation by a Flat Polynomial
Supervisor: Prof. Sourav Chakraborty, Advanced Computing and Microelectronics Unit, ISI, Kolkata

Projects

1. Lipid-Starch-Cell Identification **[Image Segmentation]**
   - To find the alternative for the fuels like Petrol, diesel and create less dependency on the crude oils
   - It is a semantic segmentation task to classify the image pixels into 4 different classes.
   - Deep Learning Architecture: Unet and Loss Function: Focal Loss
   - Performance Metric: Mean IoU Score ($\sim$93%)
2. IRM Data Labelling Analysis **[Natural Language Processing]**
   - Reduce the manual tagging efforts on data fields and automatically predict the IRM data tags
   - Algorithms Used: Random Forest (AUC: 0.75),XGBoost (AUC: 0.886),Linear SVC (AUC: 0.69),Naïve bayes (AUC: 0.6)
3. MRN News Analytics(Sentiment Analyzer for Reuters News Articles) **[Natural Language Processing]**
   - To remove the redundant data and extraction of the useful information and find the polarity score/sentiment score
   - Find out the sentiment for the given article using the VaderSentiment library
4. Spot Freight Rate Forecasting **[Time Series Analysis]**
   - To forecast the flat rates of different routes of crude vessels
   - ML algorithms: ARIMA, ARIMAX, svr and tunned svr,decision trees,random forest,knn, lasso, linear model(lm)
   - Evaluated AICc,BIC,RSE, adjusted r-squared, mape for each mentioned ml algorithms
5. GSPL Natural Gas Digitisation

- A web crawler (bot) automatically login to a GSPL website and based on the different values from the dropdown menus, it pulls the data within a given day interval and pushed that data into the MySQL database.
- Selenium is used for the web scraping and SqlAlchemy is used for the database connectivity.

6. Freight Forward Analysis (FFA)
7. Vessel Tracker
    - A lots of crudes and vessels are coming daily, so we have to track them like what would be the estimate time of arrival and what would be the temperature of a particular chemical etc.
    - All these things come in an summarised report and then an email is triggered to the business.
8. Crude Indifference Usecase(MOPS data analytics)
    - By using Marginal Value analysis and Absolute Value Analysis, top and bottom constraints are evaluated.
9. Labelling and Classification of Fake News on Social Media, Supervisor: Dr. Malay Bhattacharya
    - The development of web interface to label and classify news article shared on social media over the Internet using the help from informed citizens.
    - Multinomial Naive Bayes classifier is applied on a dataset with labels of fake or real.
10. Implementing Single and Multilayer Perceptron Model, Supervisor: Dr. Rajat De
    - Implemented the single layer perceptron model on sonar dataset which is linearly separable
    - Implemented the multilayer perceptron model for the classification of Iris flowers and the classification of handwritten characters in Mnist data.
11. Implementing Linear Regression Model, Supervisor: Dr. Swagatam Das
    - Implemented Linear Regression for the prediction of flat prices, given flat size and number of bedrooms in the flat.

Skill Set

- **Machine Learning Algorithms:** Linear/Lasso/Ridge Regression, Logistic Regression, Support Vector Machine(SVM), LDA, QDA, Naïve Bayes, KNN, Ensemble learning, Clustering, Deep Learning, NLP, Time Series Analysis, Genrative AI, Image Processing
- **Programming Languages:** C, Python, Julia, R, Matlab(Basic), Prolog
- **Libraries:** Pandas, Numpy, Matplotlib, Scikit-learn, PySpark, SqlAlchemy, TensorFlow, PyTorch
- **Databases:** MySQL, MSSQL, PostgreSQL, MongoDB
- **Tools and Framework:** Microsoft Azure, Jupyter Notebook, RStudio, Visual Studio, Git, Docker, MLflow, Streamlit, FastAPI, Tkinter
- **Environment:** Markdown, LaTeX
- **Web Technologies:** HTML, CSS, Django(Basic)
- **Operating Systems:** Linux, Windows

Extra-Curricular Activities/Achievements

- AIR 2 in JEST(Theoretical Computer Science) in 2019.
- 98.36 percentile in GATE 2018.
- 1st Prize in All India Sanskriti Gyan Pariksha from year 2002 to 2006 and 1st Prize in Computer Education in 2004.
- Attended Computer Application Course under A.I.C.P. conducted by All India Society for Electronics and Computer Technology with A+ Grade in 2003.