HARSH JAIN

M.Tech (Industrial and Management Engineering)

Institute email : harshjain20@iitk.ac.in

Personal email : jainharsh335@gmail.com

Phone : +91 7348851465

 GitHub

 LinkedIn

Academic Qualifications

Year	Degree	Institute	CPI/%
2020 - Present	M.Tech.(IME)	Indian Institute of Technology, Kanpur, U.P.	7.0*/10
2016-20	B.Tech.(Agriculture Engineering)	CAE,UAS, Raichur, Karnataka	8.672/10
2016	XII	Shree Vidhya Sagar Public School, Khategaon, M.P.	76.6%
2013	X	Pushpdeep International School, Khategaon, M.P.	82.56%

Internships

Urvija AI | Product Development Intern   (*May 17, 2021 - July 25, 2021*)

- Researched about the **business problems** in the agriculture sector and have to search and develop the solution for the same with the help of insights from the satellite data and to make **Business Requirement Document (BRD)** including the business problem and **Product Requirement Document (PRD)** including the solution for the same.
- worked on the solution to find the feasible facility location for agro-based industries; Created BRD and PRD for the same; included Avoided cost as my **North Star Metric (NSM)** and **KPIs**.Created **Needs Statement, RAID Framework, Developed Product Roadmap** and got the glimpse about the roles and responsibilities of **Product Managers**. Check my work

Key Projects

- **Credit Card Fraud Detection** *(Course Project) [Stochastic Process]*   (*March'21 - May'21*)
  Instructor: Prof. Avijit Khanra, Department of Industrial and Management Engineering, IITK
  - Used simulated data of a credit card user to train a **Hidden Markov Model** from scratch and estimated transition probabilities and emission probabilities using **Baum-Welch Algorithm** (Forward-backward algorithm)
  - sequentially predicted whether the upcoming transaction is fraudulent or not with **recall 0.81** and **F1** score **0.67**.
  - Tools Used: R(HMM). GITHUB
- **Life Expectancy Prediction** *(Course Project) [Data Mining and knowledge Discovery]*   (*October'20 - November'20*)
  Instructor: Dr. Faiz Hamid, Department of Industrial and Management Engineering, IITK
  - The objective is to predict the life expectancy by studying the 15 years supervised Data of **193 countries** with **22 attribute**.
  - **Outliers** and **null** values are treated with **winsorization** and **median imputation** technique.
  - Attributes **dropped** using **Heat Map**. Data visualization involves **scatter** and **distribution** plots.
  - **Random forest** regression is finalized for modelling with **r-squared** value of **0.95** and **4.3 MSE**. GITHUB
- **Binary Classification** *(self Project)*
  - The objective is to classify whether or not the students will be admitted. This dataset contains 400 records with 4 attributes.
  - **EDA** includes data visualization, **descriptive** analysis, **univariate, bivariate** analysis, heatmap for correlation.
  - **Decison Tree Classifier, Logistic Regression and KNN** models applied with **hypertuning** and compared based upon **Accuracy Score**. KNN is finallized with accuracy of **72.15** percent. GITHUB
- **Amazon Fine Food Reviews by NLP** *(Course Project) [Applied Machine Learning]*   (*February'21 - March'21*)
  Instructor: Dr. Veena Bansal, Department of Industrial and Management Engineering, IITK
  - Sentimental Analysis of reviews of food items from amazon.The supervised data includes 10 attributes with five lakh records.
  - **text Normalization** of raw data which includes removal of white spaces, converting to **lower** case,removal of **punctuation, lemmatization, tokenization**, removal of **stop words, vectorization** has been done.
  - Applied **logistic regression** to classify reviews as positive and negative. package used: NLTK(python) GITHUB
- **Clustering** *(Course Project) [Applied Machine Learning]*   (*March'21 - April'21*)
  Instructor: Dr. Veena Bansal, Department of Industrial and Management Engineering, IITK
  - The objective of the **unsupervised** data-set is to make **clusters** of customers of a mall with 5 attributes.
  - Examing the data set for null & duplicates, **descriptive statistics**, visualization using **pair-plot**, data analysis with **dabl**, checked for **class imbalance** and finally clustering analysis using **Kmeans**.Packages: dabl,seaborn, matplotlib GITHUB
- **Time Series Analysis** *(self Project)*
  - To predict the daily temperature in Melbourne using 10 years data. checked for Auto/serial correlation by **Acf** and **lag plots**, stationarity by **Adfuller**, seasonality or any trend by **seasonal decompose visualization**.
  - **1-step prediction** model has taken as **baseline model**, Used **RMSE** as performance measure.
  - Applied the **SARIMA** and **AR** model. Used **Grid** search and **lag vs error plots** for optimum **lag value**.
  - Finally,**SARIMA** is decided by visualizing the prediction results from all three models. GITHUB
- **Portfolio Optimization** *(Course Project) [Financial Engineering]*   (*March'21-April'21*)
  Instructor: Dr. Suman Saurabh, Dr. Shankar Prawesh, Department of Industrial and Management Engineering, IITK
  - To **gain max return** by investing in top 15 stocks chosen based upon the **Sharpe ratio**.
  - Portfolio is optimized by **Markowitz optimization** Theory. The virtual trading is done at Money Bhai with **Rs 1 crore**.
  - **Expected** return is **Rs 532415** with **Risk** involved **23.5** percent. **Actual** Return is **Rs 35,140** within a **month**. GITHUB

- **Virtual Equity Research Experience Project** *(By Quollab)* *(July'21)*
  - Linked The 3 types of financial statements namely, **Profit and loss, Balance sheet and Cashflow** with the data of ABC company for year 2016 and 2017. Created the **Revenue Model to Forecast** for the year 2018 based upon the information gain during **Analyst Call**. Got **EBIDTA** Rs 14,141.54, **closing balance** as Rs 8,943.30, **Debt-Equity Ratio** as 4.03, **Current Ratio** as 0.34, **Net Profit-Ratio** as 0.067 and some other ratios. Tools: Excel. Check my work

Positions of Responsibility

- **Treasurer in M.Tech, IME Department, IIT Kanpur** *(September'20 - Present)*

Technical Skills

- **Programming Languages and Tools:** Python, R, SQL, MS Office(Excel, Power Point, Word), Tableau
- **Libraries and Packages:** Pandas, Numpy, Matplotlib, Seaborn, SciKit Learn,Statsmodels, NLTK, Keras/tensorflow (basics)

Relevant Courses

Data Mining and Knowledge Discovery	Probability and Statistics	Financial Engineering
Applied Machine Learning	Stochastic processes and Applications	Introduction to Computing
Supply Chain Management	Operations Research	Advanced Decison Models

Achievements

- Awarded **HackerRank 5 star Gold Badge** in **SQL**
- Secured an **All India Rank of 11** in **GATE 2020** with 99.42 percentile

Extra-Curricular Activities

- NSS special camp as volunteer at Kalmala village, Raichur. (Theme- "Youth for Greenery, Soil & Water Conservation")
- Long distance Runner, Chess player, Avid Reader