# ABHISHEK SAHU

abhisahu20@iitk.ac.in; +91-7388965491

**M.TECH (Industrial & Management Engineering)** **LinkedIn:** **GitHub:**

## ACADEMIC DETAILS

| YEAR | DEGREE | INSTITUTE | PERFORMANCE |
|---|---|---|---|
| **2020-22** | **M.Tech (Industrial & Management Engineering)** | **Indian Institute of Technology, Kanpur** | **8.52* CPI** |
| 2015-19 | B. Tech (Textile Chemistry) | Uttar Pradesh Textile Technology Institute, Kanpur | 72.02% |
| 2013 | Class XII \| UP Board | Vivekananda HS School, Kanpur | 88.88 % |
| 2011 | Class X \| UP Board | Vivekananda HS School, Kanpur | 69.33% |

**Up to 2nd Sem*

## INTERNSHIP

**Data Science Intern** at **Harvesting India Pvt Ltd** (AgriFin Company) *(April'21-July'21)*

**Detecting Defects in Fruits Using Deep Learning**

**Problem Statement:** Using **Deep learning** to automate fruit detection, sorting and identification of damaged ones

**Methodology**:

- Collected 2000+ pictures of each fruit that would cover the variety, and split the data into (70/30) ratio of training and testing
- Selected a *YOLOV5* (a heavily altered version of Darknet 53) model using Image-AI to train using our collected dataset

**Deliverables:**

- Successfully created an **AI model** that can be used in fruit farming, production and packaging & providing better quality products to customers
- Initiated model deployment on the Harvesting India website, obtained 82% accuracy on the test data set

## ACADEMIC PROJECTS

**Statistical Modelling for Business Analytics**

**Telecom Customer Churn Prediction** **ongoing* *(Aug'21-Sep'21)*

- Dataset contained 7043 rows (customers) and 21 features such as" tenure", "online security", "paperless billing" etc
- Performed EDA, applied **SMOTE** to balance the data **and RFE (Recursive Feature Elimination)** to select the 15 significant features
- **Logit** and **Probit** models were used for classifying the Churn class, features were dropped based on **p-value** and **VIF**
- Logit showed better results as **accuracy** of about **79%** ,**precision** of **73.8%** and a **recall** of **62.4%**,**AUC** of **ROC curve** was **0.83**

**Applied Machine Learning**

**NYC Taxi ride pick-up prediction** *(Feb'21-Apr'21)*

- Created time interval using the UNIX timestamp, data processing using **Dask-DataFrame ,** used **K-means clustering** to break NYC regions
- Applied **Baseline Models** on 'ratio' and 'previous' values along with **Linear Regression, R-F Regression and Xg-Boost**
- Baseline Model (Exponential Moving Average) worked best on MAPE performance Matrix with **14%** on train and **13.6%** on test data set

**Data Mining**

**Fake Job Description Prediction** *(Oct'20-Nov'20)*

- Performed data pre-processing steps on the data that consists of both textual information and meta-information about the jobs
- Analyzing the text data using **NLP techniques**
- Applied **Logistic Regression, KNN, SVM, Random Forest Classifier, Sklearn's** on the trained data set
- **Logistic regression** performed well with **F1-Score 0.98** and **Precision of 97%** with accuracy around **95%**

## SELF PROJETCS

**NLP & Deep Learning**

**Amazon Sales Recommendation** *(Jan'21-Mar'21)*

- Build a **recommender system**, which recommends similar apparels items using text and image data, data obtained through **web scraping**
- Performed **deduping** text pre-processing & data cleaning actions to reduce the 180k rows & 19 columns dataset to 16k rows & 7 columns
- Applied **Bag-of-words** (BoW) , **Term Frequency-Inverse Document Frequency** (TF-IDF) & **word2vec** (W2V), W2V showed better results
- Used **CNN model** (VGG16) on visual features based on similarity offering good recommendations, 'A|B' testing can be done all models

**Power BI Dashboard**

**Sales Insights for a Computer Hardware Company** *(Jan'21-Mar'21)*

- Dataset included 150k+ datapoints of sales data, including the customers, dates, markets, products & transactions
- Performed Data cleaning and ETL (Extract, transform, load) using currency normalization and handling invalid values
- Created Key performance indicator in Power BI dashboard to get insights, which would help in making data driven decisions

## RELEVANT COURSEWORK AND SKILLS

| | |
|---|---|
| **Coursework** | **Statistical Modelling** for **Business Analyst** \| Applied **Machine Learning** \|**Data Mining** and knowledge discovery \| **Probability and Statistics** **Stochastic Process** and their Applications \| **Operation Research** for Management \| **Operation management** |
| **Skills** | **Technical Skills: Python**(Numpy,Scipy,Panda,Matplotlib, Scikit-learn) \| SQL \| Power BI \| DAX \| MS office, Excel \| Google Analytics |
| | **Professional Skills :** Product Management & Development, Statistical Analysis, Supply chain Planning |

## WORK EXPERIENCE (9 Months)

**Assistant Merchandiser, Mahavir SpinFab Ltd, Unnao** ( Safety Garments Division) *(June'19-March'20)*

- Responsible for **Product Development**, **Business Strategy** and Order Execution while being a part of a 5-member team
- Developed several **samples**, prepared design of various products and contributed to good customer service, to the existing buyer and new one
- Connected with new buyers by conducting **deep-dive analysis** various companies to assist in the competitor analysis and strategy making
- Leveraged planning knowledge to drive supply chain activities, through the creation modification & implementation of products plans

## ACHIEVEMENTS/AWARDS & CERTIFICATIONS

- Secured **AIR 6 in GATE** 2020 (TF) with 99.65 percentile
- Finished in Top 1% Students in Uttar Pradesh in 12th Board
- Selected in **"INSPIRE SCHOLARSHIP FOR HIGHER EDUCATION"** Under the Govt. of India.

## POSITIONS OF RESPONSIBILITY

**Core Team Member of Web Team, IME IIT Kanpur** *(Aug'20- Present)*

- Inter-Department Students Body to facilitate all the web related needs of students and website management of the IME department
- **Coordinated** the end-to-end process of data collection and updating relevant details of 300+ alumni of MTech IME, IIT Kanpur

**Teaching Assistant for the Course MBA 664A ( Supply Chain Management)** *(Aug'21- Present)*

- Handling of course logistics and contributed to discussions for improving the course content and delivery