SHUBHAM AWASTHI

☎ +91-8318511892 | ✉ sawasthi20@iitk.ac.in | GitHub: ⬤ | LinkedIn: 🔗

EDUCATION
*up to 2nd Semester

Degree	Institute / Board	CGPA / Percentage	Year
MTech (Industrial and Management Engineering)	**IIT Kanpur**	9.62*	2020-22
BTech (Mechanical Engineering)	**IIT -ISM Dhanbad**	8.0	2013-17
Senior Secondary Education	CBSE	92.2%	2012
Secondary Education	ICSE	87.8%	2010

PROFESSIONAL EXPERIENCE

Data Science Intern | Next Labs, Mphasis
May'21-July'21

Project-1 | MLOps- Feature Engineering Pipeline

- Developed python scripts for pipeline with functionalities such as interaction variables generation, date-time feature transformation etc.
- Developed **Custom Transformers** using scikit-learn base classes and **OOPs** concepts to serve the pipeline components.
- Integrated **Mlflow** tracking component to log artifacts, parameters, metrics and user settings of ML pipeline.

Project-2 | AWS Marketplace Solutions

- Developed predictive analytics solutions that address the No-Show problem in health-care services and reservation cancellation problem in the hospitality industry. The solutions are published on **AWS Marketplace** as: Medical Appointment No-Show Predictor & Hotel Reservation Cancellation Predictor
- Trained/developed the respective models on **Ensemble Machine Learning** algorithms, to predict the likelihood of No-Show in medical appointments and Cancellation in hotel bookings.
- Served the model as **REST** API using **flask**, followed by containerization using **Docker** and deployed the model on **AWS Sagemaker**.
- Explored, debugged and interpreted the trained complex ML models using Explainable AI python package- **SHAP**.

Sr.Analyst / Software Engineer | Capgemini India
July'17-May'18

- Responsible for developing **RESTful** Microservices using Spring Boot for solutions enabling real-time monitoring of freight containers.
- Developed and maintained data layer using ORM framework, and wrote **MySQL/JPA** Hibernate queries for **CRUD** operations.
- Implemented all the components following a test-driven methodology using **Junit** and **Mockito**.

ACADEMIC PROJECTS

Quora Question Pair Similarity | NLP Text Mining | Data Mining and Knowledge Discovery
Sept'20-Dec'20

- Developed a binary classification model to identify questions on Quora that are duplicates of the questions that have already been asked.
- Performed EDA, text normalization (**stop words removal**, **tokenization**, **stemming**) followed by advanced feature extraction using Fuzzy features.
- Analysed features by visualisation using **T-SNE and word clouds**.
- Featurized text data using **tf-idf weighted word-vectors** and created a baseline model to get the worst-case log-loss performance.
- Applied **Logistic Regression**, SVM and **XGBoost** with hyper parameter tuning and evaluated the models on precision & log-loss.

Time Series Forecasting of CO$_2$ Emissions by Electricity Generation in USA | TSMBA
Jan'21-May'21

- Performed statistical test such as **ADF, KPSS & PACF-ACF plots** to identify trend, seasonality and stationarity in data.
- Decomposed time series using **STL and SEATS,** Applied **Holt Winter's Additive Model** and **SARIMA** model for forecasting the emissions.

Predicting quality of Red Wine using Statistical Regression Models* | Statistical Modelling for Business Analytics
Jan'21-May'21

- Checked for **correlation**, **multicollinearity** and verified **OLS assumptions**.
- Identified heteroskedasticity using Breusch-Pagan test and corrected the consequences of it by using robust standard errors in regression.
- Performed **multivariate regression analysis** to predict red wine quality using 1599 data observations.

Analysing efficacy of eLearning at IIT Kanpur | Marketing Research
Jan'21-May'21

- Conducted a **Problem Identification Research** to analyse & identify salient features of eLearning at IITK.
- Collected 200+ sample data (primary) using **online surveys**, **focus groups** & **personal interviews**.
- Carried out Exploratory, Descriptive Research in **SPSS** on the data (Convenience Random Sampling).
- Analysed the data using statistical (parametric & non-parametric) tests (**One /Two Sample t-test, Chi squared test**) to verify hypotheses.

Decision Support Systems (DSS) | CADS
Jan'21-May'21

- Developed a **DSS** with user-interface and a dashboard to forecast the demand for a goods leveraging **HTML, PHP and MariaDB** on IIT-Kanpur's server.
- Designed **ER diagram** for a bank application and developed a database using MariaDB with satisfying various entity relationships and cardinalities.

COURSEWORK, SKILLS AND CERTIFICATIONS
*audited

Thesis Work	**Named Entity Recognition (NER)** Based Approach to Extract Crime Information from Online Indian Newspaper Articles and Analyze the Crime data using Statistical **Causal Inference Methods**.
Technical Skills:	**Python** | **Data Structures and Algorithms** (Basics)| **SQL** (MySQL) | ML Libraries: NumPy, Pandas, Matplotlib, **Scikit-Learn, Statsmodels, Shap** | DL Libraries (Beginner): **Keras**
Academic Courses:	**Applied Machine Learning** | **Data Mining & Knowledge Discovery** | **Probability & Statistics** | **Statistical Modelling for Business Analytics** | **Time Series Modelling for Business Analytics** | **Computer Aided Decision Systems** | **Introduction to Computing** | **Marketing Research**
Online Certifications:	**Machine Learning A-Z: Python** | **Learn SQL Course, Codecademy** | Introduction to Data Modelling with Power BI

Positions of Responsibility

Department Placement Coordinator IME | SPO, IIT K
July'21- Present

- Working as a departmental representative of SPO, IIT Kanpur to assist in successful conduction of placement drive.

Overall Student Coordinator, Srijan'16, IIT-ISM Dhanbad
Sep'15-Mar16

- Responsible for managing the promotion, sponsorship, design and technical team together for the proper functioning of the event.
- Coordinated the conduction of 30+ competitions and nine major events with a team of over hundred volunteers.
- Meticulously managed budget of INR 4 million and negotiated with lead academies of Dance, Musicals and Fine Arts to provide mentorship and other non-monetary Incentives to the outstanding cultural performers in Srijan.

Hospitality Team, Srijan'15, IIT-ISM Dhanbad
Sep'14-Mar'15

- Responsible for all managing all the logistics related to hospitality and accommodation of 200+ participants with 3 other members.

VP Membership, TMCG Club, Toastmasters International
Jan'21-Jul'21

- Monitor membership levels and strategize with the club executive committee to handle membership challenges.

Sergeant At Arms (SAA), TMCG Club, Toastmasters International
Jan'21-Jul'21

Extra-Curricular & Scholastic Achievements

- Secured **99** percentile in **GATE 2020** among the 1,37,826 applicants.
- Secured **5138** rank in **IIT-JEE Advanced'13** among the 1.26 million applicants.
- Secured **340** GEN Rank in **UP State Engineering Entrance Examination.**
- Participated in **Analytics Case Study** competition event during **Prabhandhan'20, IIT Kanpur.**
- Working as **Freelance Subject Matter Expert (Physics)** at Chegg Inc.