# Roshan Sharma

**Email-id:** roshansharma.datascience@gmail.com

**Phone Number:** 8249411154

## About Me:

**Summary:**

**I am a B.tech Computer Science Student from Lovely Professional University, Punjab Specialized in Machine Learning, Data Analysis, Data Visualization, and Deep Learning. I am a Quick Learner and aspire to become a Machine Learning Engineer or Data Scientist. I have worked on more than 100 projects on Data Science. I have a keen interest and passion for Data Science.**

**Online Platforms:**

- **Kaggle: https://www.kaggle.com/roshansharma**
  - **Kernels Expert, Rank 94**
  - **Discussions Expert, Rank 114**
- **Github: https://github.com/sharmaroshan**
  - **74 Repositories**

**Courses and Specializations:**

- **Machine Learning Specialization from Coursera**
- **Advanced Machine Learning Specialization from Coursera**
- **Data Visualization Specialization from Coursera**
- **Deep Learning Specialization from Coursera**
- **Data Analysis Specialization from Coursera**
- **Business Analysis Specialization from Coursera**

## Education :

- **South Point Public School, Patna, Bihar**

  **I have studied in South Point Public School located in suburbs of Patna provides a really serene environment to the campus. I have Studied PCM(Physics, Chemistry, and Mathematics) here. I secured 76% Marks in my CBSE Board Exams.**
- **Lovely Professional University, Jalandhar, Punjab**

  **Lovely Professional University (LPU) is a private university situated in Phagwara, Punjab, India. I am studying B.tech Computer Science. I have been a consistently good performer in all academic activities and tasks. I have a curent CGPA of 9.3. I have taken Machine Learning as my Minor Subject and Psychology as my Non technical Minor.**

## Skills:

- **Python and R**
- **Machine Learning**
- **Deep Learning**
- **Data Visualization**
- **Data Analysis**
- **Tableau**
- **Natural Language Processing**

## Projects:

- **Fraud Detection in Online Payments :**

**It is a Project based on Anamoly Detection, where I have used Machine Learning and Deep Learning Hybrid Models for Prediction and Sampling to avoid the bias in the distribution as the dataset has a distribution of 0.05:99.95 for the non-frauds and frauds, the model may turn out to be bias so it is very important to apply sampling. In this case I have applied SMOTE. I have also performed Queries, Analysis, and Visualizations for better picture formation.**

***Outcome: Able to detect the Frauds in an online Payment System.**

- **Employee Attrition Rate Analysis :**

**It is a project based on Predicting the attrition rate of the employees to help the HRs. I have used R programming for the Prediction Analysis and High Charter for better and Interactive Visualizations to better understand the picture. Ad-hoc Analysis by putting specific questions to get the most clear and effective answers.**

***Outcome: Able to predict the Attrition Rate and also understand the reason behind it.**

- **Spam Detection :**

**It is a project based on Text analytics and Classification of Spam Users from the Regular Users using Natural Language Processing, and Machine/Deep Learning. Data Visualizations, Analysis, EDA etc has been performed to analyze the situation and get the deeper insights of the problem. Using TF-IDF for feature extraction from the messages and then applying a suitable project to classify the spam users.**

***Outcome: Able to predict the Spam messages.**

- **Predicting Money Spent in Resort :**

**It is a Project based on Analytics Vidhya Hackathon Sponsored by Tech Mahindra. It is a Advanced Regression Problem where Statistics, Maths, Different Ensembing Techniques such as Stacking, Boosting, etc. I have used Advanced Machine Learning Models such as Lgboost, XgBoost, Random Forest etc to get a better accuracy. For understanding the depth of the attributes I used seaborn to visualize them with the dependent variable.**

***Outcome: It can Predict the Money spent by a visitor at the Resort.**

- **Insurance Claim Prediction :**

**It is a Data Science solved by using Machine Learning and Data Visualization. The Aim is to predic the Insurance Claim of any Visitor to the Bank. EDA, Visualizations, Analysis is performed to get the insights of the data. Different Models such as Random Forest and AdaBoost Regressors are used and metrics like RMSE Score, RMSLE Score etc are computed to see if the model works better.**

***outcome: It can Predict the Insurance Claim of any Customer.**

- **EMI Prediction :**

**It is a Data Science Problem solved using Advanced Machine Learning Techniques. Statiscal Package Scipy is used to better understand the data. Different Models have been stacked together and then boosted with lgboost and xgboost(best working models) to get the desired results.**

***Outcome: It can Predict the EMI**

## Tools and Technologies:

- **Packages: Sklearn, Mlxtend, Scipy, Numpy, Pandas, Matplotlib, Seaborn, PyTorch, Fastai, ggplot, plotly, High Charter, Ntlk, Spacy, Apriori, Kmeans, PCA, SVD, TSNE, Bubbly, networkx etc.**
- **Models: linear Regression, Random Forest, SVM, Decision Forest, K Nearest Neighbors, AdaBoost, lgBoost, XgBoost, Naive Bayes, Logistic Regression, Lasso, Ridge, ElasticNet, Auto Encoders, CNN, RNN, Resnets, VGG, etc**
- **Techniques: Regression, Classification, Recommendation, Clustering, Dimensionality Reduction, Data Visualizations, Geo-Spatial Analysis and Visualization, Anamoly Detection, Image Classification, Social Network Analysis, Market Basket Analysis, Sentiment Analysis, Ad-hoc Analysis etc.**

## Hackathons:

- **Titanic Dataset: Top 4%**

- **Blind Detection:** Top 10%
- **Aerial Cactus Identification:** Top 7%
- **Advanced House Price Predictions:** Top 14%
- **Predict Future Sales:** Top 28%
- **Don't Overfit:** Top 6%
- **Piathon-2:** Rank 1