# Ankit Sharma

**Email**: ankit1404sharma@gmail.com
**Mobile No:** +91-8826958861
**LinkedIn:** https://www.linkedin.com/in/ankit-sharma-72184494/
**GitHub:** https://github.com/Ankit-Sharma1404

## o Career Objective:

Seek to work in an environment that will challenge me further while allowing me to contribute to the continued growth and success of the organization.

## Experience Summary:

- **Having 3.5 years** of experience in Data Science & Machine Learning, development in the Health **Insurance** domain.
- Experience in **Statistical, Data Modelling, POC, Visualization and Clustering.**
- Hands-on experience in **Python, Hive, SQL, NoSQL, PySpark**.
- Hands-on experience in implementing Machine Learning regression and classification models such as **Linear regression, Logistic regression, KNN, Naïve Bayes, NLP, Decision Trees, Random Forest**, and **K-means clustering**.
- Model Optimization through **K-Fold** cross Validation and Parameter tuning through **XGBoost**, dimensionality reduction through **PCA** methods.
- Root cause analysis and **Cost Benefit Analysis** through Pareto graph chart analysis.
- Extensive experience in all phases of software development life cycle and followed agile methodology development.

## Professional Summary:

- Software Engineer in **Global Logic Technologies Pvt. Ltd. (**HealthCare Insurance Domain), Gurgaon from **July-2018 to till date (2.10 Years).**
- Software Developer in **Sellers Cure the Business Development (**Crawling**),** Gurgaon from **Jan-2018 to June-2018 (6-Months).**
- Internship in **KPMG**, Gurgaon from **July-2017 to October-2017 (3-Months).**

## Project #1:

### Project Name: Health Insurance Claim Fraud Detection
### Role: Software Engineer
### Organization: Global Logic Technologies Pvt. Ltd.
### Team Size: 12
### Client: Apax Partners

**Description:** The machine learning models that are discussed and applied on the datasets were able to identify most of the fraudulent cases with a low false positive rate i.e. with a reasonable precision. This enables loss control units to focus on new fraud scenarios and ensure that the models are adapting to identify them. Certain datasets had severe challenges around data quality, resulting in relatively poor levels of prediction.

### Responsibility:

- **Business Problem:**
  - Machine learning techniques allow for improving predictive accuracy, enabling loss control units to achieve higher coverage with low false positive rates i.e. with a reasonable precision to analyze fraud claims and fraudulent using classification and Regression Algorithms like **Naïve Bayes, KNN, SVM, Decision Trees, Random Forest**, and **K-means clustering etc** using programing language Python (pandas, numpy, scikit-learn, matplotlib, URLLib etc.) and for database **SQL** or **Hive**.
  - While building detection models, the savings from loss prevention needs to be balanced with the cost of false alerts.
  - Extracting data from **SQL, Hive or PySpark** and explaining the current trend of members on our portal using visualization techniques such as **Matplotlib** and **Seaborn.**
- **Observation:**
  - Covering-up for a situation that wasn't covered under insurance (e.g. drunk driving, performing risky acts, illegal activities etc.)

- Inflating the impact of the incident: Increasing the estimate of loss incurred either through addition of unrelated losses (faking losses) or attributing increased cost to the losses.

- **Predictive Analytics Solution:**
  - Explore various machine learning techniques to improve accuracy of detection in imbalanced samples.
  - The impact of feature engineering, feature selection and parameter tweaking are explored with the objective of achieving superior predictive performance.
  - As a procedure, the data will be split into three different segments – training, testing and cross-validation.
  - Once the dataset is obtained and cleaned, different models are tested on it
  - Based on the initial model performance, different features are engineered and re-tested
  - Once all the features are engineered, the model is built and run using different fi values and using different iteration procedures (feature selection process)
  - To improve model performance, the parameters that affect the performance are tweaked and re-tested.
- **Achievements:** Got 3 times appreciation mail sent by client for my work and on time delivery.

**Project #2:**

**Project Name: GLIPCS**
**Role: Software Engineer**
**Organization: Global Logic Technologies Pvt. Ltd.**
**Team Size: 14**
**Client: Apax Partners**
**Description: GLIPCS** is **Global Logic Internal Portal for Customer Services** having Insurance benefits from **Apax Partners**. In this portal Members can Login in to their account and see their benefits, Claims, add Dependents, Spouse and Pay for their Claims etc.
**Responsibilities:**

- **Business Problem:**
  - Data modeling to predict the likely number of claims to be filed by the customer in consecutive months and financial years using **Classification** and **Regression** Algorithm using programing language Python (pandas, numpy, scikit-learn, matplotlib, URLLib etc.) and for database **SQL**, **Hive** or **PySpark**.
- **Observation:** Modelled data to get the probable number of customers who shall retain our insurance policy.
- **Predictive Analytics Solution:**
  - Creating ad-hoc reports as per onshore requirement and based on the Problem statement Design Models are created.
  - POC from scratch which later developed as a product.
  - Coordinating with different teams regarding multiple different tasks.

- **Achievements:**
  - Received Global Logic Excellence Award and Positive feedback from Leadership group.
  - Gave training in Python, ML Algorithms and enrolled as a trainer in Global Logic for internal programs.

**Project #3:**

**Project Name: Web Crawling**
**Role: Software Developer**
**Organization: Seller Cure the Business Development**
**Team Size: 6**
**Description:** To develop Customer Management to collect, analyze and interpret raw data from various websites for existing customers.
**Responsibilities:**

- Utilize web scraping techniques to extract & organize competitor data and Modernize data streaming process, resulting in 25% redundancy reduction.
- Harvesting textual data from the web and cleaning it up.

## Project #4:

**Project Name: Migrate KPMG Portal Pages.**
**Role: Internship**
**Organization: KPMG**
**Team Size: 32**
**Description:** To define a product and develop a monitoring dashboard for KPMG Internal Portal.
**Responsibilities:**

- Developed Spam Analyzer an Intelligent spam filtering system.
- Local Search Refiner – Refines to be used for generating pre-filtered URLs should be configured and available in the productive environment.
- Developed scraper for multiple online retail websites using Regex and Beautiful Soup.
- Quality Analysis of pages on Crawled heavy scripted enabled.

## Technical Certificate:

- PG-Diploma for **CDAC (Centre for Development of Advanced Computing)**
- Certified in **Google Analytics** by Google.
- Certified in ML, Data Science, Statistics, Mathematics, Probability, and SQL from **'365DataScience'**

## Education Qualification:

| Degree | Institute | Board | Year | Marks |
|---|---|---|---|---|
| PG-Diploma | Centre for Development Advanced Computing (CDAC) | Centre for Development Advanced Computing (CDAC) | 2016 | 62% |
| B.Tech | Uttaranchal Institute of Technology (UIT, Dehradun) | Uttarakhand Technical University (UTU) | 2015 | 73% |
| 12th | St. Paul Schools | CBSE | 2010 | 55% |
| 10th | Bal Vikas Vidyalaya | CBSE | 2008 | 65% |

## Personal Details:

| | |
|---|---|
| Date of Birth | 14-Sep-1993 |
| Marital Status | Single |
| Email | ankit1404sharma@gmail.com |
| Mobile | +91-8826958861 |
| Passport No. | N2525125 |
| Skype | a270744106c51de0 |
| Permanent Address | Moh. Lashkariganj, Sasaram, Rohtas, Bihar-821115 |
| Current Address | Gali No.-4, Block-K, Mahipalpur, New Delhi-110037 |

.