ASHUTOSH PUNIA   **Contact No +919996209054**

Email ID: ashutoshpunia001@gmail.com
Github: https://github.com/ATP001
Kaggle: https://www.kaggle.com/atp001/account
Linkedin: https://www.linkedin.com/in/ashutosh-punia-bba855b1/

Summary

- Predictive modeling using Python, Machine Learning Techniques like XGBoost, Linear & Logistic Regression, Decision Trees, Random Forests, PCA, Naive Bayes Classifier, Clustering & Segmentation, Text Mining & Classification
- Image Classification using Tensor flow, Keras
- Machine Learning expertise – Regression, Classification, Feature Engineering, Data Scrapping (See Projects) , Manipulation and visualisation (e.g Matplotlib, Seaborn, Plotly)
- Python –scikit Learn, Numpy, Pandas, Matplotlib, SciPy, Tensorflow, Keras, Plotly

Skill Set

Languages Tools Framework Statistical & Modeling Techniques	- Proficient in performing data exploration, preparation and analysis on massive datasets using Python - Familiar with SQL queries Hive,Pig - Hadoop Ecosystem, especially Hive data warehouse infrastructure and HQL - Regression Techniques – Linear | Logistic - Cluster Analysis | Decision Trees | Random Forest| Support Vector Machines Hypothesis Testing | ANOVA | Chi Square | Correlation

Technical Dexterity

- Wiley certified Big data Analyst
- Oracle Certified Java Programmer
- IABAC Certified in Data-science foundation
- Machine Learning proficient
- Worked on Hadoop cluster running CDH4.4
- Worked with highly unstructured and semi structured data
- Extracted the data from cvs files into HDFS using Hive / Pig
- Implemented extracting real time Twitter data into HDFS
- Implemented extracting real time Twitter data into Anaconda using Tweepy, TextBlob, Seaborn, Sklearn (Python packages)
- Worked on ANN (Image Classifications with Tensorflow and Keras)
- Exploratory Data Analysis
- Deep learning enthusiast

May 2016 – Dec 2017 :- DataMites Pvt Ltd (Intern in Data Science)

- Improving performance of machine learning algorithm by parameter tuning
- Studying the performance of different models by n-fold cross validation
- Mathematical analysis of gradient descent technique in neural networks and how can we improve it. Reporting all the findings of the experiment and and provide conclusions that can help in better data modelling in future.
- Worked on python in Anaconda distribution.

`

Project : 1

Email Classification

We Used different Natural Language Processing techniques for classifying email into various classes. For this task, we created TF-IDF values for the words present in the training emails in Python. On top of it, we used Linear Support Vector Machine (SVM) and Logistic Regression models and got 85% accuracy

Project : 2

Analytics Case study

Case Study 1: Analyze Marketing Campaigns of Magazine Publisher (Wiley Certification project)	
Operating System	Ubuntu
Languages/Tech	Hadoop , Hive ,Pig
Databases	Hive
Tools Used	Tableau

Project Description:

Email Marketing Campaign

The project is to Analyze email data fetched from a Marketing email campaign of a magazine publisher. The High level design is fetching the user behavior on the campaign email send to them and how they responded on this. Data is in the form of csv file and its then moved to HDFS. All data transformation is done with some pre-calculations. Reports are displayed on the dashboard CTOR by any field provided by the client. And business rules are implemented for the report generation. e.g. number of head count for per household. The exploratory data analysis and information collected can help in predictions.

Roles and Responsibilities:
- Data preparation methods using PIG and then some Data queries using Hive.
- Queries are to perform basic exploratory analysis on the data available.
- End objective is to find which audience is interested in my offer or not with the email send.

Rubixe (Jan 2018 – Jan 2019)

Consultant Data scientist
Project : 3

To Build a Predictive Model and find out the sales of each product at a particular Retail Store.

- Performed extensive analysis in order to understand the properties of products and stores which play a key role in increasing sales through graphical representation.
- Explore and visualize the data, build simple, base models for benchmark.
- Conceptualized feature engineering and leveraging the business insights from Case Review like (Count of Outlet Identifiers, Count of Item Identifiers and Outlet Years etc.)

`

and setup a robust/thorough validation framework consistent with the evaluation metrics.
- Performed methodical experiments for the selection of optimum Tree based parameters.
- Used various ML Models like Linear, Decision Tree but Random Forest outperforms other Models with its lowest RMSE value.

Project : 4

To Identify the Customers Segments and Target those customers who are Eligible for Home Loan.

- Explored data after understanding the problem statement, evaluation metric and data dictionary. Tried various forms of summaries and visuals, after which I began the modelling.
- Used feature engineering and Create Variables like (Total Income, Income by loan etc.) in order to order to build successful machine learning solutions.
- Being a binary classification problem, the algorithm used like Logistic Regression, Random Forest, GBM and XGBoost came into play.
- Used 5-Fold Cross Validation Technique and Tune the features in order to improve the predictive power of the model.
- Evaluate Metric is accuracy i.e. percentage of loan approval you correctly predict.

Project : 5

Twitter stream A lot of companies monitor mentions from their customers on the internet to react to the negative ones ASAP. For example, T-Mobile and Verizon need to respond to negative tweets fast and find out what the problem is and how they can solve it.

We can make this project using a convenient Twitter API and sentiment analysis algorithms to detect such tweets in the whole stream.

ML problem: sentiment analysis
Algorithms: sentiment analysis
Technologies: nltk, spaCy
References: Twilert, Tweetreach, Tweepy

Industrial training

- Underwent Training in NIIT For Java, C,C++,
- Oracle Certified Java programmer
- Network Security And Management
- Attended Enablement Session organized by M/s IBM Bluemix .
- Training From Jigsaw Academy for Wiley Big Data Analytics including Tableau
- IABAC Certified Data-science Foundation

Educational Qualification

Qualification	Board/University
B.Tech. (CSE)	MMU, Mullana, 2016
Senior Secondary	C.B.S.E
Secondary	C.B.S.E

`