



Keshari Gupta

(1118241)

ML Engineer

[TATA CONSULTANCY SERVICES]

Address: [WFH] Sukhdehra,
Ghazipur, Uttar Pradesh
Zip code: 233231

Mobile: +91-7752851242

Email: Keshari.g@tcs.com

SUMMARY

Having 4.9 years of experiences in python automation and data science to gain progressive idea and insight from large structured datasets, leveraging mathematics and applied statistics with visualization.

USED ALGORITHMS

➤ Machine Learning

- Logistic Regression
- Linear Regression
- Support vector Machine
- KNN
- MultinomialNB, GaussianNB
- Random Forest
- XGBoost

➤ Deep Learning

- Deep Neural Network(MLP)
- Convolution Neural Network (Transfer Learning)
- RNN(LSTM,GRU)
- Encoder-decoder and Attention Model (in progress)

Experience

Project: Continuous Manufacturing automation of downstream purification

Role: Research and development [Noida VI STP, 01-Feb-2019 to Till now]

Description: Working with IIT Delhi on continuous manufacturing research for downstream purification. Ph.D. students of IIT Delhi are researching on downstream purification and we are providing the solution to automate the process of downstream.

Roles and Responsibility:

- Providing the solution to automate the BioSMB, Acoustic wave separator and HPLC.
- Implemented the python framework using design pattern (Abstract factory method, command pattern and singleton design pattern) which is fetching instruction from PLC DB and performing the same on the machine like open pumps, valves switch on/off
- Integrating MATLAB model with Python so MATLAB model can help in continuous train by predict the flow of pumps ,acoustic powers ,aeration rate and agitation rate .
- Helps to make communication with machine (PLC) and software of machines.

Project: Credit Card Lead Prediction

Role: Data scientist [AnalyticsVidhya-A-Thon May 2021]

Description: Happy Customer Bank wants to cross sell its credit cards to its existing customers. The bank has identified a set of customers that are eligible for taking these credit cards. The bank is looking for your help in identifying customers that could show higher intent towards a recommended credit card.

Roles and Responsibility:

- Removed duplicate rows and perform EDA to figure out useful features.
- Data was "Imbalanced" so tried with oversampling (SMOTE and Random Sampling).
- Preprocessed data and added new features (Feature Engineering) for better Model.
- Applied One Hot Encoding for categorical features and trained various models where XGBClassifier worked well and got 0.835 AUC score on CV data and 0.832 on test data.
- I scored 351 rank on Analyticsvidhya Job Thon's leaderboard.
- **Code:** https://github.com/KeshariGupta-ML/analyticsvidhya_job_a_thon_may21

Case Study: Personalized Medicine: Redefining Cancer Treatment

Role: Data scientist [Aug-2020]

Objective: It is multi-class classification problem and objective is to classify the given genetic variations/mutations based on evidence from text-based clinical literature.

Roles and Responsibility:

- Applied preprocessing and removed HTML tags, stop words, punctuation and expanded contractions etc.
- Performed EDA and figure out which feature is more important using univariate analysis and that feature is stable or not.
- Converted categorical variable into numerical using response coding technique and text feature to vector using TFIDF and TFIDF-W2V techniques.
- Used GridSearchCV for tuning of hyper-parameter with multi-class log loss. And found best model with 0.89 log loss on test data.
- **Code:** <https://github.com/KeshariGupta-ML/Genetic-Variants-effect-prediction>

EDUCATION	Experience
<p>MCA: <i>JSS Academic of Technical Education, Noida (UP) (Batch-2013-2016) – 77.5%</i></p> <p>BCA: <i>TERI PG College, Ghazipur (UP) (Batch-2010-2013) - 77.5%</i></p> <p>Intermediate (12th): <i>Govt. Inter College, Ghazipur (UP Board) (Apr 2009 – May 2010) – 68.80%</i></p>	<p>Case Study: Predict Physician Segment for Drug marketing</p> <p>Role: Data scientist [June-2021]</p> <p>Business Problem: Axtria Company had launched a drug; however, some physicians are yet to prescribe it for the first time. A key client stakeholder has reached out to a Decision Sciences Principal in Axtria for help to identify potential physicians who are most likely to start prescribing the drug in the next quarter in order to channelize the marketing efforts more effectively while targeting potential physicians.</p> <p>Roles and Responsibility:</p> <ul style="list-style-type: none"> Formulate business problem into Machine learning problem. Clean data and perform univariate, bivariate and multivariate analysis on data and figure out which features are more contributing for predicting target. Along with EDA , also checked multicollinearity in data using VIF method and also plotted Pearson correlation and Spearman correlation with heat map. After Analysis and preprocessing, converted categorical data to vector using one-hot-encoding and trained multiple model like Logistic regression, SVM, KNN, Random Forest, LightGBM to get best model. Best model was LightGBM which gave minimum log loss (Train log loss: 0.4, CV log loss: 0.8, Test log loss: 0.8) Saved lightGBMclassifier and deploy this model on heroku using streamlit application. Code: https://github.com/KeshariGupta-ML/Predict-Physician-Segment
USED LIBRARIES	
<ul style="list-style-type: none"> ➤ SkLearn,LightGBM ➤ Tensorflow,Keras ➤ Pandas, numpy ➤ NLTK, Glove W2V ➤ Matplotlib ➤ networkx 	
CERTIFICATION	
<ul style="list-style-type: none"> ➤ Applied AI course (in progress and completed 26/32 assignments) ➤ Improving Deep Neural Networks: Hyper parameter tuning, Regularization and Optimization on Coursera ➤ Neural Networks and Deep Learning on Coursera ➤ Machine Learning on Coursera 	<p>Case Study: Amazon Fine Food Review</p> <p>Role: Data scientist [July-2021]</p> <p>Problem Statement: The Amazon Fine Food Reviews dataset consists of reviews of fine foods from Amazon. And objective is for a Given a review, determine whether the review is positive (Rating of 4 or 5) or negative (rating of 1 or 2).</p> <p>Roles and Responsibility:</p> <ul style="list-style-type: none"> Preprocessing of text like removing HTML tags, stopwords, punctuation, performing stemming and expanding contractions etc. Perform exploratory data analysis like word cloud, heat map etc for positive and negative data points. Generate TF-IDF features using TfidfVectorizer and after getting TF-IDF scores, convert each question to a weighted average of word2vec vectors by these scores. Build and evaluate the model <ul style="list-style-type: none"> Naive Bayes Logistic Regression - with L1 and L2 regularization Linear SVM RBF Kernel SVM
Social Link	
<ul style="list-style-type: none"> https://github.com/KeshariGupta-ML https://www.linkedin.com/in/keshari-nandan-gupta-211151111/ ML portfolio Keshari Gupta (kesharigupta-ml.github.io) 	