

Professional Summary

- Aspiring Data scientist with advanced knowledge of Python, Statistics, Machine learning and Deep learning. I have a strong foundation and experience in data cleaning, Exploratory data analysis, building statistical and text classification models while pursuing PG Data Science and Engineering (DSE) from Great Learning. And, effectively communicating the finding in the form of a story to both IT leaders and business stakeholders.
- GitHub: <https://www.github.com/impraveenchand>
- LinkedIn: <https://www.linkedin.com/in/praveen-chand-kakarla/>

ACADEMIC PROFILE

Name of the Program	College Name	Month-Year	Percentage/ GPA
Post Graduate Program in Data Science (PGP-DSE)	Great Lakes Institute of Management	May-2019 to Oct-2019	Excellent
BE/B. Tech	V.R. Siddhartha Engineering college (Autonomous)	July-2015 to April-2019	7.54
12 th Std	Narayana Junior College	June-2013 to April-2015	9.52
10 th Std	Ravindra Bharathi public school	June-2012 to April-2013	9.3

ACADEMIC PROJECTS

Natural Language processing (NLP)	Using Natural Language Processing to Build A Spam Filter For Text Messages <ul style="list-style-type: none"> RandomForestClassifier and GradientBoostingClassifier were used to determine a text message was spam or ham. With 100% precision Random Forest is predicting the False positive accurately. But the Gradient boost model got 88.9% precision score. Spam filter – optimize for precision. In this case False positives are very costly. So, we choose Random Forest Model.
	Multi-Class Text Classification to Predict Happiness Source by using sklearn and Keras libraries. <ul style="list-style-type: none"> Based on the happy moment statement you have to predict the category of happiness. Text Preprocessed using Scikit-learn's and keras(tokenizer) libraries. Built a pipeline to vectorize the text data, then train and fit a model. After that evaluate the performance of the model on the test data. Selected the model which has high F1_score. Linear Support Vector classifier model has the highest f1_score (93 %) out of all the machine learning models. By using LSTM as the neurons for the hidden layers, I built the deep neural network model and got 88.7% accuracy on the test data. The LinearSVC model is performing better than the LSTMs model on the test data. So, the best model to predict the happiness source is LinearSVC.

Machine Learning (ML)	Libraries used: NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn, SciPy, Stats Models, Itertools, keras and imblearn.
	Salary Prediction for AMCAT Test Takers: <ul style="list-style-type: none"> ▪ Considering both R² score (53.1%) and RMSE (97,532) the best model the dataset shows is Ridge Regression. ▪ However, the high RMSE scores are pointing towards wrongful prediction of the target variable. ▪ Hence, we will be dividing Salary into multiple classes and try to build a classification model over it. ▪ Treated imbalanced data with SMOTE. ▪ XG Boost Classifier is performing better than the other models after appropriate hyper-parameter tuning with 84% precision and 84% recall. Credit Risk Analysis: <ul style="list-style-type: none"> ▪ Used Logistic Regression to predict whether a person will default the loan or not ▪ Found positive correlation between credit history and loan approval and also between applicant income and loan amount ▪ Created a predictive model which had an accuracy of 95.05%.

CORE SKILLS

Technical

- Python
- Tableau
- MySQL
- Exploratory Data Analysis
- Math & Statistics
- Story-telling and Data Visualization.

Transferable

- Data driven decisions
- Business Strategy
- Communication
- Problem Solving
- Time Management

CERTIFICATIONS AND ACHIEVEMENTS

- Participated in a **hackathon hosted by Hackerearth to Predict the source of happiness** and got a 92% score on the Leaderboard
- Completed **"NLP with python for Machine Learning Essential training"** course on Lynda.
- Completed **"Applied Machine Learning: Foundations"** course on Lynda.
- Accomplished **" Applied Machine Learning: Algorithms "** course on Lynda.
- Finished **" Intermediate Python for Data Science "** course on DataCamp.
- Completed **"Intro to SQL for Data Science"** Course on Datacamp.