

CHEERAG VERMA

DATA SCIENCE
ENGINEER

CONTACT

+91-98188-74-188
cheeragverma188@gmail.com
www.cheeragverma.com

OBJECTIVE

To have a rewarding and fulfilling career in which I can satisfy my thirst for knowledge and practice the lessons learnt to deliver quality work to the best of my abilities, for the betterment of my employer and in return expect to be enriched in terms of knowledge and experience.

SUMMARY

- 2.5+ years of professional experience in Application Development using Python, Machine learning, Pandas, Numpy, Matplotlib and Django.
- Strong problem solving skills using Data Structures and Algorithms in python.
- Hands on experience and exposure on Kaggle with Exploratory Data Analysis, Data Preparation, Data Modelling and Feature Engineering techniques.
- Strong ability in building models on large datasets by utilising advanced Statistical Modelling, Machine Learning techniques.
- Good Hands on Unix and deployment.
- Extensive implementation and contribution to the Open-Source community GitHub and medium.
- Strong communication skills and ability to thrive in a fast paced, team-oriented environment while delivering quality work by tight deadlines.

SKILL SUMMARY

LANGUAGE/SCRIPTS	PYTHON/SHELL SCRIPTING
Machine learning technology	Scikit-learn, Numpy, Pandas, Matplotlib, Python, Data Structure and Algorithms
Business Domain	Financial Domain
Repository Management	Git, BitBucket, SVN
Strengths	Statistics, Maths and Data Structure, Python

PROFESSIONAL EXPERIENCE

1.CitiRisk NAV Automation

IRIS Software

Jul 2019- Till date

Client : CitiBank,US

Description: CitiBank is a well established enterprise in the Financial domain & one of its verticals deals in Hedge Fund. Currently, It receives thousands of reports from different hedge fund houses with different financial reports in PDF format. Company invests a huge chunk of money in extracting information from PDF reports manually which is time taking and error prone.Created an NLP Engine that will parse monthly Funds Business performance reports received from various clients in various formats like PDF, Excel or mails. The tabular structure present in the PDF can be extracted using statistical techniques.

Role: Developer

Contributions:

- Development of data pipeline for the analytics platform .
 - Involved in requirements gathering, analysis, design & code implementation.
 - Data preprocessing to cleanse and improve the quality of data.
 - Participated in troubleshooting, bug fixing and production moves.
 - Responsible for writing Workflow, Move to Production, Scheduling and Previewing Modules.
 - End to End deployment in higher environment and code maintenance.
-
- Skills : Statistical Analysis of Data, Data Visualisation, Predictive Modelling, Machine Learning,
 - Deployment.Environment/Tools: Spyder, Jupyter Notebook, Microsoft Visual Studio Code, Jupyter Notebook, Git, Python.

Highlights : Extracting the tabular data with high accuracy from PDF applicable on all PDF files through exploiting the statistical distribution of data pattern present. Currently, no open source tools are available in the market which extracts the Tabular data from PDF Files and Images with such high accuracy and performances.

2. Contract Analyser

IRIS Software

Sep 2018-Till Date

Client : CitiBank,US

Description: The financial enterprises follow the legal framework to perform the business activities whenever two or more entities are involved. To comply with the legal obligations, analysts perform the task of checking the legal obligations in the contract manually, which is normally hundreds of pages. Contract Analyser assists financial analysts to reduce their efforts by extracting the relevant data based on their queries without going through the details of lengthy contracts.

Role: Developer

Contribution:

- Developed an code to extract Entities from tabular structure data from contract.

3. Inferential & Descriptive Data Analysis For The Credit Card Business

IRIS Software

Client : CitiBank,US

June 2018 - Sep 2018

Description: The client deals in the credit card business and has a user base in millions in the North American region. Daily, hundreds and thousands of transactions are carried out by the users of varied backgrounds from different categories for various reasons. The project aims at drawing the inference from the user data and their transactional pattern to make best decisions in the interest of the business through the data.

Role: System Engineer

Contribution:

- Developing the data pipeline to ingest the data for data analysis purposes from different sources.
- Statistical analysis of data through Normal, Poisson & Erlang probability distribution.
- Drawing the inference by data visualisation in Matplotlib library Identification of groups in the data through the clustering technique KMEANS.
- Detection of outlier activities in the data like, fraud transactions. Predicting the behaviour of the user base

Skills : Statistical Analysis of Data Using Scipy, Data Visualisation Using Matplotlib, Hierarchical & KMeans Clustering, Logistic Regression, Time series Data Analysis Using Stochastic Processes

Education

University/School	Course/Class	Percentage	Year of Passing
Guru Gobind Singh Indraprastha University - Northern India Engineering College	B.Tech.- IT	78%	2018
St. Francis De Sales School	XII	85%	2014
St. Francis De Sales School	X	95%	2012

Certifications

- Certification in Data Structure and Python from Coding Ninjas
(<https://ninjasfiles.s3.amazonaws.com/certificate775900cad6807c47046022742998ec5898eaId.pdf>)
- Certification in Python from NPTEL- Joy of computing from IIT-Roper, secured top 1% and featured on their website. (<https://nptel.ac.in/noc/courses/noc19/SEM2/noc19-cs41/>)
- Successfully Completed Coursera Machine learning course
- Certification in data science machine learning from udemy.