

Sandeep Kumar Are

Mobile: +91 7989520772

E-mail: sandeepkumar.are16@gmail.com

LinkedIn

Personal website

CAREER OBJECTIVE

A proactive and fast learning individual seeking an opportunity to work as a dynamic data analyst utilizing analytical & methodical skills and relevant expertise to help the company achieve business goals while sticking to vision, mission and values.

PROFESSIONAL SUMMARY

- Total **6.0+** years' experience in IT software Development field and in that **3 +** years in **Python, Machine Learning** with large data sets of **structured data, data validation, predictive modeling and data visualization**.
- Experience on **NLP** with **Deep Learning** Models.
- Experience on Tensor Flow 2.0 and Keras
- **Good Experience** with **Machine Learning Supervised / Unsupervised Learning algorithms**.
- **Good Experience** with **Probability and Statistics for Machine learning**.
- Adept in statistical programming language **Python**.
- Good Experience with **NLP NLTK/Spacy** Libraries.
- Proficient in managing entire data science project life cycle and actively involved in all the phases including data acquisition, **data cleaning**, features scaling, **statistical modeling**.
- I have experience on **Web scraping** data with **Python-Selenium**.
- I have Experience in **Python – MongoDB**.
- Extensive experience in **Text Analytics**, generating **data visualizations** using statistical programming (**Python**).
- Good Experience on Text Pre-Processing with **Bag of Words (BoW), TF-IDF, Word2Vec, and Avg word2vec**.
- Good Experience on Topic Modeling LDA and NMF Techniques.
- Good Experience with **SQL Server 2008/2012/2014/2017**.
- Good industry knowledge, analytical & problem-solving skills and ability to work well with in a team as well as an individual.
- Highly creative, committed, intellectually curious, business savvy with good communication and interpersonal skills.

Deep Learning Concepts:

- **Neural Networks: ANN, CNN and RNN**
- **Tensor Flow 2.0 With Keras.**

Machine Learning concepts:

- **Dimensionality Reduction:** Principal Component Analysis, TSNE.
- **Prediction Analytics.** Simple Linear Regression.
- **Classification techniques:** Logistic Regression, K-Nearest-Neighbor Methods.
- **Tree-based Methods:** Regression Trees, Classification Trees.

- **Ensemble Methods:** Bagging, Random Forests, Boosting.
- **Support Vector Machine:** Maximum Marginal Classifier, Support Vector Classifier.
- **Natural Language Processing with Deep Learning.**
- **Time Series Analysis with Python.**

EDUCATIONAL QUALIFICATION

- ★ B.E Electronics & Communications engineering **JNTU University, India.** 2012

ROLES & RESPONSIBILITIES

- Review the dataset and keen understanding of target output to be solved.
- Taking care of dataset with all necessary steps like missing values, checking correlation.
- Performed Data Cleaning, features scaling, features engineering, feature standardization with Count vectorization /TF-IDF vectorization NLP techniques.
- Worked with different techniques from **NLTK / SPACY**.
- Involved in data visualization with Matplotlib, seaborn techniques.
- Evaluated models using Cross Validation.
- Addressed over fitting by implementing of the algorithm regularization methods.
- Used Principal Component Analysis and t-SNE in feature engineering to analyze high dimensional data.
- After Exploratory data analysis, analyze which model is fit for the data.
- Analyze the results with difference measurement techniques Confusion matrix, Classification reports, Accuracy, Precision, Recall etc.
- I have Good Experience on TIME SERIES ANALYSIS.

WORK EXPERIENCE

Projects:

Working as a Software Engineer (AI, ML, NLP and Python) in Svobodha Infinity Private Limited (SAVART) from February 2020 to Till Date.

Project Name: Question and Answer Model.

Built QA model where it fetches the relevant data from the closed domain data. This is achieved by just placing the search bar in a UI portal and backend operations is done by running **BERT** Model based on the keywords which are placed in search box.

Roles and responsibilities:

- I have involved scrap the data from various sites, API's and some more resources using Python Selenium.
- Involved in working with Python-MongoDB various tasks.
- Involved in Topic Modeling with NLP.
- Worked with team on BERT Model.
- I have experience in working with Flask-Python for creation of UI.

Project Name: Vantage QUANT.

Vantage QUANT completely related Financial domain Project. Worked with Stock market related Quantitative DATA. Involved in DATA Cleansing, DATA Analysis, Features selections, Predicting Quantitative outputs.

Roles and responsibilities:

- I have involved in what are approaches are gives the best results.
- Involved in Data cleansing, Data Cleaning, Data Analysis.
- Involved in Feature selection Models.
- Involved in Forecasting Models selections to predictions.

Worked as a Software Engineer in Netpeach Technologies Pvt Ltd from February 2015 to January 31st2020.

Project Name: Uncover Autism

Project Description:

The goal of the project is to help society mitigate the risk of autism in new born babies by identifying more reliable causes of autism. The project will be exploring existing research, and available information with feedback(surveys) from the patient population to define more reliable causes. The purpose of this study is to provide the healthcare industry with any findings discovered during analysis.

Tools and Technologies:

- Environments Anaconda Jupyter Notebook, NLTK tools.
- Numerical and Visualization libraries NumPy, Pandas, Matplotlib, Seaborn.
- Machine Learning libraries Scikit- learn.

Projects:

Loan Predictions:

Given historical data on loans given out with information on whether or not the borrower defaulted (charge-off), can we build a model that can predict whether or not a borrower will pay back their loan? This way in the future when we get a new potential customer, we can assess whether they are likely to pay back the loan.

*This is which I have completed while doing TensorFlow 2.0 exercise from UDEMY Certifications.

Worked with: Tensor Flow 2.0 , Feed Forward Neural Network(ANN), Kera's with TF.

<https://colab.research.google.com/drive/1HbwYVE5erBLsqdwO1xqqYYRX1GqB5IFL?usp=sharing>

Predict the Image on CIFAR-10 Dataset:

From CIFAR-10 Data set we have 10 different kind of Images. So through TF2.0-CNN how can we predict that what evaluating the Image is predicted right or wrong.

https://colab.research.google.com/drive/1a8jEazm8GL05A-DDvAh5NaOO_ND6epQs#scrollTo=XwWqSFfkPULp

My-Portfolio(Personal Website) - <https://sandeepkumar16nlp.github.io/sandeepmyportfolion.github.io/>

Certifications:

NLP with Python for Machine Learning Essential Training

Statistics Foundation1.

Python for Time Series Data Analysis.