

Sandeep Kumar Are

Mobile: +91 7989520772

E-mail: sandeepkumar.are16@gmail.com

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

- Worked as Software Engineer in Netpeach Technologies Pvt Ltd.
- Total 4.6 years' experience in IT software Development field and in that 1+ years in **Machine Learning** with large data sets of **structured data, data validation, predictive modeling and data visualization**.
- Adept in statistical programming language Python.
- I have good experience in Django with PostgreSQL.
- 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.
- Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data.
- 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 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.

TECHNICAL SKILLS:

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.
- **Python Programming.**
- **django and PostgreSQL.**

WORK EXPERIENCE

Worked as a Software Engineer in Netpeach Technologies Pvt Ltd from February 2015 to August 2019.

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.
- Implemented LDA and LSA Topic modeling techniques in Data sets.
- Performed Data Cleaning, features scaling, features engineering using pandas and NumPy packages
- Used clustering technique K-Means to identify outliers and to classify unlabeled data.
- Evaluated models using Cross Validation.
- Addressed over fitting by implementing of the algorithm regularization methods.
- Used Principal Component Analysis in feature engineering to analyze high dimensional data.
- Performed Multinomial Logistic Regression, Decision Tree, SVM to classify package is going to deliver on time for the new route.

Projects:

Uncover Autism

Role: Worked as Data Analyst.

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.

Roles and Responsibilities:

- Data definition, collecting the data as per data definition.
- Extract the data using Python and involved in Exploratory Data analysis on different data sets using Pandas, Matplotlib and seaborn for visualization to understand what data we have.
- Involved in Data cleaning with NLTK. Data Preprocessing with deduplication, text pre-processing NLP.
- Performed Data Cleaning, using pandas and NumPy packages in python.
- Involved in Dimensionality reduction techniques PCA and TSNE.
- Analyze the data and apply best Classification or regression algorithms.
- Involved in working with Train, Cross validation and Test data sets.
- I have worked with KNN classification and Logistic regression techniques.
- Analyzing the data using required software (Python) by applying required mathematics and statistics.
- Pass the clean data to data science team.

Tools and Technologies:

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

Synchronoss Intranet Portal

Role: Team Lead and SharePoint developer(SharePoint Online).

Client Name: Synchronoss Technologies Inc.

Project Description:

Synchronoss wished to create a comprehensive enterprise-wide intranet portal to enable its employees to access & share information, collaborate their day to day work activities, etc.

The client is now able to connect their globally dispersed workforce more effectively, breaking down the barriers to firm-wide communication with Easy-To- Use Global Online Intranet Solution, leveraging Office 365, Results in Collaboration, Organization and Standardization. I have much experience on creating workflows using SharePoint Designer 2013 and worked on creating .aspx pages, master pages as page layouts.

Technology: SharePoint Online, JavaScript, jQuery, JSON.

SharePoint Migration and SharePoint Support(2010/2013).

Role: Team Lead, SharePoint Admin and Developer.

Client Name: Haemonetics, The Blood Management company.

Project Description:

Haemonetics has been leveraging Microsoft SharePoint portal technologies for document management, team collaboration, corporate intranet portal, and much more. The company needed to upgrade its SharePoint environment from 2010 to 2013 to take the advantage of new features of SharePoint 2013 such as

Technology: SharePoint Enterprises Server 2013,

Role: Administrator, Developer.

Team size: 3

Technical highlights:

- SharePoint Enterprise Server 2013.
- SharePoint Designer 2013
- SharePoint Management Shell Power Shell scripts.

Samples:

1. Exploratory Data Analysis(EDA) on Iris Data set.

I have implemented Exploratory data analysis on Iris Data set.

https://github.com/SandeepKumarAre/Datascience/blob/master/EDA_IrisData/AI_EDA1.ipynb

2. Dimensionality reduction technique: PCS and t-SNE:

a. Here I used MNIST data set.

b. https://github.com/SandeepKumarAre/Datascience/blob/master/Dimensional_reduction%20Techniques_PCA_TSNE.ipynb

c. https://github.com/SandeepKumarAre/Datascience/blob/master/t-SNE_Dimensionality%20Reduction%20Technique.ipynb

3. Text Pre-processing: BoW, TF-IDF, Word2Vec and AvgWord2Vec on Amazon Food Reviews

a. https://github.com/SandeepKumarAre/Datascience/blob/master/TextProcessing_BO_W_TFIDF_W2V.ipynb

Blog: <https://sandy1604.blogspot.com>