**Haasini.Yerneni**

**Mobile: 7842719889**

**E-Mail:** haasiniyerneni99@gmail.com

**Summary: Data Science Engineer** involved in full lifecycle Total 3 years 6 months of experience in **Data Science** to include Requirements, Preprocessing – Data Cleaning, Modeling – Training and Testing, Model Tuning and Deployment. I am Critical thinker has strong **analytical, math** and **statistics** knowledge to turn data into information, information into insight and insight into business decisions.

* Ability to analyze large datasets to discover trends and patterns.
* Ample knowledge data **pre-processing, profiling, cleansing, validation, reduction** and **transformation.**
* Expertise in **Future Selection, Future Engineering, Imbalanced /Balanced** data handling and **Exploratory Data Analysis**
* Good Experience in Working Libraries like **NLTK** and **Spacy**.
* Experienced on **Natural Language Processing** (**NLP**), **Sentiment Analysis, Lemmatization, Stemming, Stop words, IDF, TF – IDF.**
* Experience in Word Embedding Models like **Word2Vec, Doc2vec (Gensim) and Bag of Words (BOW).**
* Very Good Experience in NLP Preprocessing Activities like **Preprocessing from the Text** and **Extract Numerical Features from the Text**
* Working knowledge and Experience on **Web Scraping**.
* **Machine Learning** – Experience in building predictive models using Regression Techniques like **Association Rule Mining**, **Decision Tree**, **Random Forest**- **Bagging, Boosting, Clustering - KNN, K-Means**, **Naive Bayes** and **SVM**.
* Knowledge on **Recommendation Systems** – **Collaborative and Content based Filtering**.
* **Deep Learning** – knowledge on **Simple Neural Networks**, **Deep Neural Networks, ANN, CNN and RNN.**
* **Artificial Intelligence** - Knowledge on **Face and Eye Detection,** Displaying the **Images with Open Computer Vision.**
* **Deploying the Applications using Flask (REST API) and Django and cloud Environment like AWS and Microsoft Azure.**
* Ample development experience in **Python 3.x, Spyder, Jupyter, Numpy, Matplotlib, Scikit, Keras, Theano, Tensor flow and R.**
* Sound experience in working with **Parameters, Actions, Filters, Legends, Measures, Dimensions, Functions, Generating Dashboard Reports with Tableau.**

**Educational Summary:**

* **B.Tech** from MGIT-E.C.E(2011 to 2015) Affiliation of JNTU Hyderabad.

**Professional Experience**:

* Working asa **Data Scientist** at **Accenture**

**Technical Skills:**

|  |  |
| --- | --- |
| **Operating Systems** | Windows 95/98/2K/NT, UNIX |
| **BI Tools** | Tableau 8.x, 9.x |
| **Scripting** | Python 3.x, Pandas, Keras, Seaborn, TensorFlow, Nltk, Re |
| **Statistics** | Descriptive, Predictive, Inferential |
| **Languages** | R – Studio 3.1, Visual Studio 2010 |
| **Database** | SQL Server, Oracle, MS Access |
| **Packages** | MS-Office 2010 |
| **Project Management Tool** | Jira 8.1, HPALM |
| **Version Control** | Git |

**Experience:**

**Data Scientist at Accenture**

**Project: Healthcare Analytics**

**Description:** This project deals with trauma cases coming to the hospital for treatment. The main objective is to find out the relationship and association between injury types. By using these hidden relations, associations & patterns doctors/physicians recommend different tests and scanning.

**Responsibilities:**

* Cleaned & transformed the data to do required analysis
* Performed **chi-square test analysis** to check test’s of independence
* Used "**Apriori**" and **FP Growth** algorithm **to mine the strong association rules(ARM)**
* **Used supervised modeling techniques using Python**
* Built multiple models to predict injury types & cause of death by using **“Scikitlearn”package**
* Performed Classification using **Deep Learning, Artificial Neural Networks**
* Compared classification models with Logistic, Decision trees, Random Forests, & SVM techniques
* Also performed Regression analysis
* Handled over plotting by using certain ways with **Matplot for Explanatory Analysis.**

**Data Scientist at Accenture**

**Project: Paraphrase Detection**

**Description:** Paraphrase detection, which means analyzing sentences that are semantically identical. We propose to detect the semantic similarity between two texts of the same language to establish the similarity. To find related sentences written in natural language is complex for various applications, like text summarization, plagiarism detection, information retrieval and question answering system etc. Realizing this gravity, we examine in particular how to mark the challenges with detecting paraphrases using Natural Language Processing Techniques.

**Environment:** Python 3.x, Jupyter, NLTK, SPACY, Beautiful soup

**Responsibilities:**

* Worked with NLTK library for NLP data processing and finding the patterns.
* Worked on building statistical models to understand language at the level of words, sentences, conversations.
* Used various NLP techniques for text representation, semantic extraction techniques etc.
* Used effective text representations to transform natural language into useful features.
* Applied Modern quantitative techniques in NLP.
* Applied text representation techniques (such as n-grams, bag of words, sentiment analysis etc.).
* Applied Text canonicalization, Chunking, Coreference resolution, Collocation Extraction and also syntactic and semantic processing Collaborate with engineering and product development teams.

**Data Scientist at Accenture**

**Project: E-Commerce Review Prediction, Fashionista, US.**

**Description:** This is a Women’s Clothing E-Commerce dataset revolving around the reviews written by customers. Its nine supportive features offer a great environment to parse out the text through its multiple dimensions. Because this is real commercial data, it has been anonymized, and references to the company in the review text and body have been replaced with “retailer”.

**Environment:** Python 3.x, pandas, Numpy, Re, Nltk, Spacy

**Responsibilities:**

* Developed python code for text pre-processing and normalization techniques, such as Language.
* Worked with noisy and/or unstructured textual data (e.g. CVs and Job Descriptions).
* Identification, tokenization, Sentence Breaking, POS tagging, Chunking
* Sentence Chaining
* Creating the feature sets and labels.
* Creating the word2vec, doc2 vec.
* Expertise in processing, evaluating and utilizing training, testing and validation data.
* Experience with libraries such as Natural Language Toolkit (NLTK), OpenC.V,NLP

**Project Name & Client: Telecommunication Analytics &SINGTEL OPTUS, UK**

**Description:** This project deals with voice connectivity and call drop problems. The main objective is to find out the reasons for call drops and voice connectivity problems. Built a classification regression model to predict a call drop

**Responsibilities**:

* Used **Seaborn, Matplotlib for explanatory analysis**
* Used data imputation techniques for missing data
* Performed **multi-collinearity analysis** to find out and handle highly correlating independent variables
* Find out the important variables causing call drops
* Performed reduction of variable dimensionality by using **scaling and PCA**
* Built a model to predict call drops by using machine learning algorithms
* Created ensemble models with boosting algorithms
* Used **confusion matrix** to interpret the output
* Handled over fitting in predictive models.

**Project: Customer Behavior with Data Analytics**

**Description:** The project focused on analyzing the company’s external customer’s behavior to achieve improved business performance through growing its customer base and increasing the revenue.

**Responsibilities:**

* Developed statistical and machine learning models that mined, analyzed and turned data into meaningful insights that are now being used to grow their customer base, and revenue in the Coffee shops
* Developed and refined complex marketing mix statistical models and worked with diverse functional groups
* Participated in all the stages in the modeling process, from collecting, verifying, & cleaning data to visualizing model results, presenting results, and making client recommendations
* Used **KMeans** to develop customer segments using unsupervised learning techniques. The clusters helped business simplified complex patterns to manageable sets that helped set strategic and tactical objectives pertaining to customer retention, loyalty, spend and acquisition