SCS SASTHA KARTHIK

Data Scientist

Address Nellore, AP, 524002
Phone 08886875050
E-mail scskarthiksastha@gmail.com
LinkedIn https://www.linkedin.com/in/karthik-sastha-723a391
WWW https://github.com/karthiksastha

IBM certified Data science professional with 3+ years of work experience in various industrious & Analytical projects. Familiar with gathering, cleaning and organizing data for use by technical and non-technical personnel. Leveraged comprehensive technical acumen and problem-solving skills to create strategies and plans for Python development projects. Excelled as an individual contributor who works well in cross-functional settings and interfaces effectively with customers. In-depth analytical experience and engagement with key stakeholders across various business users.

Skills

- Python Libraries: Pandas, Numpy, Matplotlib, Statsmodels, Seaborn, Scikit-Learn etc
- R
- Machine Learning capabilities - Classification (Naïve Bayes, KNN, SVM, Decision Tree, Random Forest). Regression Models (Linear, Multiple, Logistic), Hierarchical, Non-Hierarchical Clustering, Text Mining using NLP, and Association Rules.
- Database Language - SQL
- Data Reporting Tools -Tableau
- Statistical Methods - Predictive Analysis, Exploratory Data Analysis, Inferential Statistics, Hypothesis Testing.
- Deep-learning models - ANNs, CNNs, LSTM, Transfer Learning
- Deep learning frameworks such as Tensorflow & Keras

Education

Jun 2013 - Apr 2017 **Mechanical Engineering**
Sree Venkateshwara Engineering College - Andhra Pradesh

Work History

Aug 2017 - **Quality Assurance Engineer**

Jan 2020 *Prime Meiden Limited(Power transformers), SEZ, nellore, Andhra Pradesh*

- Wrote, edited and updated project manuals and technical documentation used by quality assurance team.
- Maintained amiable relationships with suppliers through difficult quality control issues, supporting healthy business interactions.
- Determined root cause of deviations and non-conforming results and implemented appropriate corrective and preventive actions throughout product development process.
- Led process improvement projects to help operations meet and exceed quality standards and reduce costs.
- Compiled and distributed weekly feedback to team leaders and managers to improve service time and quality while increasing productivity.
- Collected and analyzed activity data and initiated, developed and recommended improvements to systems, processes and procedures to increase productivity and reduce cost.
- Provided analytical, planning and coordination support on projects as assigned, reviewing, interpreting, analyzing and illustrating data to stimulate and support enlightened decision making.
- Halted production line in case of major non-compliance of specifications, standards or quality discovered during inspection process.

Jan 2020 - Current

Data Scientist Intern

Innodatatics, Hyderabad, Telangana

- Applied statistical and algebraic techniques to interpret key points from gathered data.
- Applied statistical methods to solve specific problems in Industries and managed assigned deliverables related to study.
- Collaborated with internal stakeholders, identifying and gathering analytical requirements for customer, product and projects needs.
- Identified and documented project constraints, assumptions, business impacts, risks and scope exclusions.
- Researched, designed and implemented machine learning applications to solve business problems affecting users.
- Transformed raw data to conform to assumptions of machine learning algorithm.
- Identified new problem areas and researched technical details to build innovative products and solutions.
- Implemented and evaluated artificial intelligence and machine learning algorithms and neural networks for diverse industries.
- Developed custom database objects, stored procedures and delivered application support.
- Built databases and table structures following architecture methodology for web applications.

Data Science Projects

1. <u>Bank Customer churn analysis</u>

· **Project goal**: The goal of this project is to build a machine learning model to predict the probability that a loan will charge off.

· **Data info**: The dataset contains all available data for more than 800,000 consumer loans issued by a large US peer-to-peer lending company where in we use only data available to investors , including information about the borrower before loan was funded

· **Approach**: CRISP-DM methodologies - Understanding business problem,Data collection,Data cleansing, EDA, Feature engineering,Machine learning supervised techniques (logistic regression,random forest & k nearest neighbours),machine learning primer, model evaluation & accuracy measures

2. <u>Predictive maintenance</u>

· **Business Problem :** A well established beverage firm is facing unexpected production break down issues which in turn impacting their productivity

· **Project Goal:** The overall goal is to reduce the production breakdown and suggest preventive maintenance schedule of plant machinery for a beverage industry

· **Objectives:**

· Extract real-time data to enable predictive maintenance

· Monitor machine performance for accurate diagnostics and predict the results

· **Approach**: CRISP-DM methodologies - Understanding business problem,Data collection,Data cleansing, EDA, Feature engineering,Forecasting techniques & LSTM,machine learning primer, model evaluation & accuracy measures

3. <u>Household Power Consumption Prediction</u>

· **Business Problem :** Power outage accidents will cause huge economic loss to the social economy. Therefore, it is very important to predict power consumption. With the rise of smart electricity meters and the wide adoption of electricity generation technology like solar panels, there is a wealth of electricity usage data available.

· **Project Goal:** Given that power consumption data for the previous months, we have to predict the power consumption for future.

· **Approach**: CRISP-DM methodologies - Understanding business problem,Data collection,Data cleansing, EDA, Feature engineering,Forecasting techniques & LSTM,machine learning primer, model evaluation & accuracy measures

4. <u>Customer segmentation:</u>

· **Business Problem :**The bank marketing team would like to launch a targeted marketing ad campaign that is tailored to specific group of customers. For this campaign to be successful, the bank has to divide its customers into at least 3 distinctive groups .This process is called marketing segmentation.

· **Project Goal:** Leverage AI/ML model for maximizing marketing campaign conversion rate.

· **Approach**: CRISP-DM methodologies - Understanding business problem,Data collection,Data cleansing, EDA, Feature engineering,unsupervised machine learning(K means,elbow method, dimension reduction-pca),visualize the clustering results

5. <u>Sentiment analysis on movie reviews</u>:

· **Project Goal:** Determination of customer satisfaction based on reviews left on services ,

classify tweets & predict them as positive or negative

· **Approach**: Text cleaning is done ,Text data is converted into numerical data, ML model is trained , applying ML algorithms - logistic, svm , Test and Evaluate the model, interpretation of the result, Implementation of BERT model for sentiment analysis.

6. <u>Email spam filter</u>:

· **Project goal**: To determine whether the text received is spam or ham using NLP techniques and ML algorithms

- **Approach:** Text cleaning is done ,Text data is converted into numerical data, ML model is trained , applying ML algorithms - logistic, svm , Test and Evaluate the model, interpretation of the result.

7. Lung disease detection:

· **Project Goal:** Detecting different kind of lung disease based on chest x-ray images provided in data

· **Data info :** Chest x-ray images and they consists of 2 diseases pneumonia and Normal, and data is available as train and test data set

· **Approach:** Importing keras, tensor flow , VGG16,image data generator, libraries,resize the image,flattening,dropping 1st and last layers of transfer learning, CNN model building,compile the model,import images,fit the model,plot the loss and accuracy,saving the file,read the image and predict the output image

8. Human body parts and Object detection:

· Face and eye detection from image and videos

· Car and pedestrian detection

· **Methods and Algorithms:** Haar cascades and open cv libraries

· **Approach**: Text cleaning is ne ,Text data is converted into numerical data, ML model is trained , applying ML algorithms - logistic, svm , Test and Evaluate the model, interpretation of the result Evaluate the model, interpretation of the result

Certifications

- IBM Certification in machine learning
- IBM Certification for python in data science
- IBM Certification for Deep Learning
- SQL for Data Science
- Mini tab for Six Sigma
- Lean management from simpli learn
- ASNT NDT Level II Certification - MT,UT,PT,RGT,VT
- Tableau certified from 360digitmg india
- Credit Risk Modelling from Udemy
- Unsupervised Machine Learning for Customer Market Segmentation from coursera