KAMALESH PRADHAN

DATA SCIENTIST

CONTACT

9883697040

KP360D@GMAIL.COM

15/1 A.K COLONY, 1ST MAIN,
TAVAREKERE MAIN ROAD,
BANGALORE–560029, INDIA

PROFILE

Hands-on 6+ years' experience in solving business problem using applied Machine Learning – Proficient in using ML algorithms like, Logistic Regression, Random Forest, SVM, Xg-boost, clustering etc. and time series forecasting techniques like ARIMA / ARIMAX / SARIMAX etc.

2+ years' experience in Natural Language processing - Text Extraction, Text cleaning, sentiment analysis, Text classification etc.

Strong programming skills in Python, R, SAS.

Have successfully delivered analytics projects for clients across industry - Oil & Gas, Logistics, Pharma, Automobile and Non-profit organizations.

SKILLS

- TREE ENSEMBLES (RANDOM FOREST, XGBOOST)
- LOGISTIC REGRESSION/LINEAR
- NLP, BERT, TENSORFLOW,PYTORCH
- TOPIC MODELING
- SVM
- CLUSTERING (K MEANS)
- TIME SERIES (ARIMA/ARIMAX/ XG-BOOST)
- DESCRIPTIVE/DATA VISUALIZATION

TOOLS

- PYTHON, R, SAS
- AZURE DATABRICKS
- SQL(BASIC)
- TABLEAU

EXPERIENCE

Bridge Tree Consultancy:

Nov 2014 – Aug 2016

Accenture

Aug 2016 – Till Date

Working for various clients to support their business through advance analytics model building using ML algorithms like Logistic Regression, Random Forest, SVM, Xg-boost along with NLP and a few statistical techniques like regression analysis, time series analysis and optimization. Also have experience in process automation like from data download to mailing to client with resulted output and business outcome driving.

Projects:

Employee Talent Profile:

Developed talent profile with the help of extracted employee skill, objective, and managers feedback from PDF in Python and with the help of Pdf minor.

A. Text Extraction

Extracted text from PDF

- Text Extraction is done with pdf minor and put them in structured using NLP.
- Extracted 4 different types of pdf of different structure from 2009-2019.
- PDFs are converted into XML format. An XML (Extended Mark Up Language) provides coordinates of each characters positioning in the PDF. This help to track and collect characters – combining them to string and finally to sentences, creating the desired extracted sentences.

EDUCATION

MSc in Statistics

2011-2013

University of Calcutta, India

BSc in Statistics

2008-2011

Ramakrishna Mission Residential College, Narendrapur, University of Calcutta, India

B. Theme tagging and Classification

Tag correct theme to sentences and classify

- Tag correct theme to the sentences through semantic matching and compute theme proficiency using WMD(Word Movers Distance) with the help of pre-trained word2vec model) based on ground truth.
- Classify into strength and development with the help of sentiment analysis using SVM/Logistic-Regression.

Order to Cash Intelligent Platform

Order to cash process, where trying to identify the invoices which will be late for payment and the invoices which have high chances to be disputed. We have created one product to target all the logistic companies which are having similar type of business.

A. Late payment

- Implemented SVM to do the Classification
- Found out 6 Features to which has been used in the model.

B. Invoice Default Reason Identification

- Extracted text from invoice image using Pytesseract / OCR and compared against customers feedback to find out default reason and direct to invoice generating section to correct rectify the invoice.

Attrition Model

Developed model to identify key drivers of attrition and predict employees at risk of attrition using machine learning techniques like Logistic Regression and Random Forest in Python.

- Drivers/Features characteristics helped to identify risk probability – selected best features based on two models and combined to get a best set of features.

Customer Selection - Automobile Parts Company

Select prospects look-a-like of customers as directed by demo profile by ranking -

- Define Trade Area from the Store Details provided by Client
- Select prospects look-a-like of customers
- Final Selection by ranking to meet the target count for each store and create report using Random Forest

Diversity Predictions

Time Series Forecasting Model –

- Designed and developed a forecasting engine to forecasts diversity and provide recommendations to improve it.
- Set up an automated process by developing 1000 + models for different workforce segments and selected the best model with minimum MAPE/AIC using ARIMA/ARIMAX model.

Funnel Analyzer -

Process Automation from data download to visualize -

- data ingestion (using SFTP) and to decrypt get key details from AWS
- Data pre-processing, KPI creation using SQL database
- Scoring based target selection
- finally ingesting the data into Tableau to create descriptive / predictive Tableau dashboard.