ROSHNI KOLI
DATA SCIENTIST

✉ roshni2511.koli@gmail.com
📞 9594972519
📍 Kalyan West - 421301

in roshni-koli
○ RoshniKoli

SUMMARY

Data Science Practitioner with 5 years of overall experience and 3 years of data science experience. I have worked on deep neural networks like image caption generation, word vector debiasing, training a custom dataset on YOLO. I have hands on experience in developing machine learning models like Regression, Classification, Unsupervised methods and Bayesian inference. Typically I use Python and its libraries pandas, matplotlib, numpy, scipy for machine learning and keras, numpy, tensorflow, gensim for deep learning. I also have hands-on experience in SQL for data manipulation.

SKILLS

TECHNICAL SKILLS AND TOOLS: Python, gensim, Tensorflow, Keras, Scikit Learn, PyTorch, matplotlib, SQL, Clustering , Regression, Classification, Predictive Modelling
DATA SCIENCE AND MACHINE LEARNING: Analytics, Data Modelling, Exploratory Data Analysis, Data Visualisation, Deep Learning, Natural Language Processing
ENVIRONMENTS: Windows 10, AWS Linux EC2, AWS S3

PROJECTS

DEBIASING WORD VECTORS
Aug. 2020 to Aug. 2020

Word Vectors tend to keep all the social biases (like Gender bias, Age bias, etc) as there are in the dataset they are trained on.
Debiasing or removing such biases from the vectors using vector difference between the word vectors and the direction of the bias.

PREDICT USER ADOPTION
Sept. 2020 to Sept. 2020

Given a dataset about user engagement with an online product, predict the user adoption.
Clean the input dataset. Generate the output parameter 'adoption' using user history data.
Use Random Forest Classifier to classify the users and generate predictions.

IMAGE CAPTION GENERATOR
June 2020 to Sept. 2020

Create automated image caption generator using VGG16 and LSTM network.
Extract image features for the flickr30k dataset using pre-loaded keras VGG16.
Clean the captions, tokenize them and create input sequences.
Develop model using addition layer for combining image features and LSTM.
Use BLEU scores and evaluate model.
Use the model with best BLEU scores and generate captions.

BOUNDING BOX PREDICTIONS - YOLO DARKNET
Apr. 2020 to June 2020

Use Darknet YOLOv3 framework to train a custom dataset of Fashion images.
The trained model should accurately predict Bounding box for each class in the image.
The dataset has 2 lakh training images belonging to 13 classes.
Perform EDA to :
Convert each image to YOLO format
Process all images and create corresponding YOLO input files.
Create GUI using PyQt and connect it to YOLO to visualize results in real-time.

CUSTOMER SEGMENTATION
Mar. 2020 to Mar. 2020

Segment the customer database into clusters using K-means.
The dataset contains information about customers transactions and offers. A separate dataset consists of product information.
Compare the number of clusters obtained using elbow method and silhouette scores.

SVM, NAIVE BAYES AND ENSEMBLE METHODS
Feb. 2020 to Mar. 2020

Predicting Movie Ratings from Reviews Using Naive Bayes
Customer Segmentation Using Clustering

PREDICT GENDER GIVEN HEIGHT AND WEIGHT
Feb. 2020 to Feb. 2020

Given data about weights and heights of people, predict their gender.
Use scatterplot on the output variable to understand the distribution and dependence on the input variables.
Create a Logistic Regression model and use K-fold cross-validation to tune the hyperparameters.

CREDIT DEFAULT
Nov. 2019 to Nov. 2019

Predicting credit card default problem forTechgig online competition using Logistic Regression.
Clean the dataset and fill in missing values.
Create visualisations to understand variables and their distributions.
Encode string variables using OneHotEncoder.
Create the model and evaluate it against the validation dataset.

REGRESSION TECHNIQUES
Jan. 2020 to Jan. 2020

Use the Boston Housing dataset to compare Linear Regression models built by selecting different variables.
Compared the models based on F-statistic and R-squared statistic.

RECOMMENDER SYSTEMS
Jan. 2015 to June 2016

Create a recommender system for cross-selling VAS product for a leading Telecom company by
using Collaborative filtering and Logical Filtering in SQL.
Use collaborative filtering and audio analytics to create curated list of caller tunes and services for Telecom users.
Use customer segmentation to find relevant segments for setting up call times.
Using customer recharge data to understand user expenditure patterns and pitch in relevant mobile products.
Develop a predictive model to recommend relevant apps based on user's influence network.
Create roadmap for product development by understand tech trends and driving innovation by interacting with the sales team.

ROSHNI KOLI
DATA SCIENTIST

EMPLOYMENT

ACCENTURE SERVICES PVT. LTD.
Associate Software Engineer
Bangalore, India
Oct. 2011 to Aug. 2013

-Trained in SAP PM, SAP SD, SAP BI.
-Worked as a SAP PM Support Engineer for Freeport McMoran Mining.
-Helped the project manager with ad hoc event management and project management work.

IKEN PERSONICS
Data Analyst
Ghansoli, Navi Mumbai, India
Jan. 2015 to June 2016

Create a recommender system for cross-selling VAS product for a leading Telecom company
using Collaborative filtering and Logical Filtering in SQL.
Use collaborative filtering and audio analytics to create curated list of caller tunes and services for Telecom users.
Use customer segmentation to find relevant segments for setting up call times.
Using customer recharge data to understand user expenditure patterns and pitch in relevant mobile products.
Develop a predictive model to recommend relevant apps based on user's influence network.
Create roadmap for product development by understand tech trends and driving innovation by interacting with the sales team.

WIPRO IT SERVICES
Senior Executive -Automation
Kanjurmarg, Mumbai, India
July 2016 to Feb. 2017

- Create chat bot services prototype for IT service-desk.
- Use banking and service-desk data to build automated solutions for common user problem
- Led 2 member QlikView team and collaborated with the client to build new reporting dashboard in QlikView
- The new dashboard saved 2 hour per day team effort of the reporting team
- Gather & analyze Business requirements pertaining to various automation products.
- Create product roadmap and strategically evolve the product to suit the changing business needs by
incorporating changing consumer behavior.
- Prepare presentation and demonstrations for client engagement and user training.
- Client interfacing and expectation management to cater to the day to day business operation needs. work
with cross-functional teams and interact with all levels
- Analyze data and derive meaningful insights on user behavior, and use the insight for product evolution.

PRICEWATERHOUSE COOPERS
Senior Consultant Data Science
Mumbai
Dec. 2020 to Current

For a pharmaceutical company, create prescriptive analysis models to help improve manufacturing yield of pellets.
- Create a variety of features based on different datasets for modelling
- Reduce number of features using Lasso Regression
- Convert the dataset into classification and regression use-cases
- Use Boruta to select best features and tune Random Forest and XGB classifiers
- Use grid search CV for Hyperparameter tuning
- Suggest best ranges for processes and parameters using SHAP plots
- Streamline data engineering pipelines

EDUCATION

Thadomal Shahani Engineering College
June 2007 to May 2011
BE Computer Science 2011

Udemy
Feb. 2018 to Aug. 2018
Machine Learning

Coursera
June 2020 to Sept. 2021
Deep Learning with Andrew Ng

Springboard
Nov. 2019 to Sept. 2020

VOLUNTEERING

YOUTH FOR SEVA · Dance and Art Teacher
Mar. 2012 to Aug. 2013
Bangalore, India

Prepare the students from underprivileged schools for extra-curricular activities organized centrally by Youth for Seva.

CRY · Student Mentor
Apr. 2016 to Feb. 2017
Kalyan

Help the children of you assigned schools to complete education by understanding their and their teacher's concerns and providing solutions.