

## SUMMARY

Data Science Practitioner with 4 years of overall experience and 2 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

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.

Aug. 2020 to Aug. 2020

### PREDICT USER ADOPTION

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.

Sept. 2020 to Sept. 2020

### IMAGE CAPTION GENERATOR

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.

June 2020 to Sept. 2020

### BOUNDING BOX PREDICTIONS - YOLO DARKNET

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.

Apr. 2020 to June 2020

### CUSTOMER SEGMENTATION

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.

Mar. 2020 to Mar. 2020

### SVM, NAIVE BAYES AND ENSEMBLE METHODS

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

Feb. 2020 to Mar. 2020

### PREDICT GENDER GIVEN HEIGHT AND WEIGHT

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.

Feb. 2020 to Feb. 2020

### CREDIT DEFAULT

Predicting credit card default problem for Techgig 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.

Nov. 2019 to Nov. 2019

### REGRESSION TECHNIQUES

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.

Jan. 2020 to Jan. 2020

### RECOMMENDER SYSTEMS

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.

Jan. 2015 to June 2016

# ROSHNI KOLI

## DATA SCIENTIST

## EMPLOYMENT

### SPRINGBOARD

#### Student

Mumbai, India  
Nov. 2019 to Current

Machine Learning and Deep Learning techniques.

Projects list:

Caption Generator  
Bounding Box detection  
Predict Movie ratings  
Predict User adoption  
Predict gender  
Predict housing prices  
Customer segmentation

### SABBATICAL AND MATERNITY

#### Upskill and self-learning

Indianapolis, IN , USA  
Feb. 2017 to Nov. 2019

Used online resources to skill up in Python and learn machine learning algorithms .

Learn about Regression, SVM, Decision Trees, Ensemble methods, Random Forests

Unsupervised Learning Techniques - Clustering.

Projects -

Predict credit default

### 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.

### 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.

### 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.

## EDUCATION

### Thadomal Shahani Engineering College

BE Computer Science 2011

June 2007 to May 2011

### Udemy

Machine Learning

Feb. 2018 to Aug. 2018

### Coursera

Deep Learning with Andrew Ng

June 2020 to Current

## VOLUNTEERING

### YOUTH FOR SEVA · Dance and Art Teacher

Bangalore, India

Mar. 2012 to Aug. 2013

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

### CRY · Student Mentor

Kalyan

Apr. 2016 to Feb. 2017

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