

ACADEMIC PROFILE /DISTINCTIONS

Year	Examination/Degree	School / University	CGPA/%
2019	M.Tech Ocean Engineering	Indian Institute of Technology Bombay	8.32/10.00
2016	B.Tech Civil engineering	National Institute of Technology Calicut	7.75/10.00
2012	CBSE	SAPS Anakkal	92.4
2010	CBSE	Warwin school Vaikom	9.6

Skilled in Machine Learning, Statistical Analysis, Programming and its application in real life problems.

Work experience

Ramboll	Jul'19-Aug'20
Roles and Responsibility :	
<ul style="list-style-type: none"> Created a machine learning model to predict the reinforcement indices of structures Analyzed the effect of structural parameters using exploratory data analysis on reinforcement indices 	

Computer Proficiency/Skills

- Programming:**
C, C++, Python, R, MySQL.
- Softwares:**
MATLAB, SPSS, Minitab, MS-Excel, MS-Project
- Familiar Libraries:**
Python- NumPy, Pandas, Scikit- Learn, Plotly, NLTK
R- Dplyr, Ggplot-2, Caret, Lubridate, Forecast
MYSQL- Select, Order by, Group by, Limit, Create Table, Update Table,
- Frameworks|Tools:** TensorFlow, Keras
- Coding Platforms:** Google Colab, Jupyter, Anaconda, R Studio, Notebook++, Bigquery
- Hackathon platform:** Analytics Vidhya, Kaggle, Hackerrank, Hackerearth, MachineHack

PROJECTS

(M.Tech course project)	<i>Time Series Analysis and study of impact of tidal condition and river discharge on Depth averaged velocity in Cochin port trust navigation channel</i>	Aug'18-May-19
--------------------------------	--	----------------------

Overview	<ul style="list-style-type: none"> Developed a time series model for depth averaged velocity using data of 4 years from 2015 to 2018. Used correlation matrix to find correlation among hydrodynamic factors and Depth averaged velocity Outliers are removed and missing values are imputed by mean values. Developed a statistical model with Multiple Linear Regression and Artificial Neural Network of 1 hidden layer. Analysed the important hydrodynamic factors using exploratory data analysis to understand its behaviour with depth averaged velocity. Analysed and developed a statistical model for tidal and river flows to understand its effect on depth averaged velocity.
Achievements	<ul style="list-style-type: none"> SARIMA model is found to be best time series forecasting model with an accuracy of 74%. Improved accuracy from 68 % in linear regression to 75 % in artificial neural network. Adding an extra feature of previous day velocity increased accuracy in both multiple linear regression and artificial neural network. Found that up to 63% of velocity during Monsoon and 79% of velocity during pre monsoon in Cochin port channel is caused by river discharge and tidal inflow respectively

Kaggle	<i>MNIST Fashion Image Classification of 10 classes with Convolution Neural Network with Keras framework (Deep Learning)</i>	Nov' 19
---------------	---	----------------

Overview	<ul style="list-style-type: none"> Dataset is collected from MNIST of Zalando's article images—consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, associated with a label from 10 classes of different fashion apparels. Keras with tensorflow as backend is used for building neural network models. Neural Network of 2 layer with RELU and SOFTMAX as activation function is developed. One hot encoding is done as data preprocessing. Convolution Neural Network model is developed of different layers and with and without padding. RELU and SOFTMAX is used as activation functions. Optimizer used is ADAM and loss used is Categorical Cross Entropy. Models are trained with different batch size and epochs to reduce overfitting and time consumed.
Achievements	<ul style="list-style-type: none"> Accuracy is used for model evaluation. Neural Network with 2 layers has test loss of 0.317 and accuracy of 88.78 % CNN with 3 layers of convolution and without padding has test loss of .21 and accuracy of 92.54 % CNN with 5 layers of convolution and with padding has test loss of .177 and an accuracy of 93.79 %.

Independent		Spam and ham classification using NLP	
Overview	<ul style="list-style-type: none"> Dataset has 5586 total observations with 2 different features of email texts and labels as ham and spam. Out of 5586 of total texts messages 746 are spam and 4822 are ham. Platform for this exercise is in python NLTK library was used for implementing data preprocessing through Natural Language Processing Data preprocessing is done using removing punctuation, tokenization, removing stopwords, stemming and lemmatizing Vectorization is done for converting text data to numerical data with bag of words, n-gram and TF-IDF. Feature engineering is done by creating extra features as body length and percentage of punctuations. Machine learning algorithm Random Forest is used for predictive modelling using vectorized data and other extra features and gridsearch with Cross Validation. 		
Achievements	<ul style="list-style-type: none"> Random Forest gives an accuracy of 97.7 % with recall of .826 and precision of 1 From confusion matrix 965 hams and 123 spams are correctly classified. 0 hams were incorrectly predicted as spams and 26 spams were incorrectly predicted as hams. 		
AnalyticsVidhya		Recommendation engine for movie ratings	
Overview	<ul style="list-style-type: none"> Collected data from MovieLens collected by GroupLens Research Project at the University of Minnesota. Datasets consists of 100,000 ratings from 1-5 from 943 users on 1682 movies and demographic information of users. Used item-item and user-user collaborative filtering method for filtering the data Used Python as programming language Used Turicreate library to build simple popularity and collaborative filtering model. 		Dec'18
Achievements	<ul style="list-style-type: none"> Able to rank movies according to collaborative filtering model. 		
Kaggle		Housing price prediction model	
Overview	<ul style="list-style-type: none"> The dataset is the prices and features of residential houses sold from 2006 to 2010 in Ames, Iowa. The training data has 1460 observations and 80 explanatory variables and the test file has 1459 observations and 79 explanatory variables. Some of the variables are Sale price, Lot area, Neighborhood, Year built, etc. And Sale price is the target variable which is the property's sales price in dollars. To understand the departure from normality distribution plot with histogram is plotted for continuous variables. 		Jan'19
Achievements	<ul style="list-style-type: none"> I used LASSO and gradient boosting regression method for developing the model. I used room mean square logarithm error as the measure of error rate. Compared RMSLE for LASSO and Gradient Boost model My model has LASSO score of .1115 and GBoost score of .1167 Average base model score was .1089 RSLME was .069 and accuracy score was .969 		
COURSES/MOOCs			
<ul style="list-style-type: none"> Machine Learning from Coursera, authorized by Stanford University : Linear Regression, Logistic regression, Support Vector machine, K Means clustering, K Nearest Neighbor, Artificial Neural Networks and anomaly detection. Python for Data science and AI from Coursera 			