

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"><li>Created a machine learning model to predict the reinforcement indices of structures</li><li>Analyzed the effect of structural parameters using exploratory data analysis on reinforcement indices</li></ul>			
Computer Proficiency/Skills			
<ul style="list-style-type: none"><li><b>Programming:</b> C, C++, Python, R, MySQL.</li><li><b>Softwares:</b> MATLAB, SPSS, Minitab, MS-Excel, MS-Project</li><li><b>Familiar Libraries:</b> <b>Python-</b> NumPy, Pandas, Scikit- Learn, Plotly, NLTK <b>R-</b> Dplyr, Ggplot-2, Caret, Lubridate, Forecast <b>MYSQL-</b> Select, Order by, Group by, Limit, Create Table, Update Table,</li><li><b>Frameworks Tools:</b> TensorFlow, Keras</li><li><b>Coding Platforms:</b> Google Colab, Jupyter, Anaconda, R Studio, Notebook++, Bigquery</li><li><b>Hackathon platform:</b> Analytics Vidhya, Kaggle, Hackerrank, Hackerearth, MachineHack</li></ul>			
PROJECTS			
(M.Tech course project)	<b>Time Series Analysis and study of impact of tidal condition and river discharge on Depth averaged velocity in Cochin port trust navigation channel</b>		Aug'18-May-19
Overview	<ul style="list-style-type: none"><li>Developed a <b>time series model</b> for depth averaged velocity using data of 4 years from 2015to 2018.</li><li>Used correlation matrix to find correlation among hydrodynamic factors and Depth averaged velocity</li><li>Outliers are removed and missing values are imputed by mean values.</li><li>Developed a <b>statistical model</b> with <b>Multiple Linear Regression</b> and <b>Artificial Neural Network</b> of 1 hidden layer.</li><li>Analysed the important hydrodynamic factors factors using exploratory data analysis to understand its behaviour with depth averaged velocity.</li><li>Analysed and developed a statistical model for tidal and river flows to understand its effect on depth averaged velocity.</li></ul>		
Achievements	<ul style="list-style-type: none"><li><b>SARIMA</b> model is found to be best time series forecasting model with an accuracy of 74%.</li><li>Improved accuracy from <b>68 %</b> in linear regression to <b>75 %</b> in artificial neural network.</li><li>Adding an extra feature of previous day velocity increased accuracy in both multiple linear regression and artificial neural network.</li><li>Found that up to <b>63%</b> of velocity during Monsoon and <b>79%</b> of velocity during pre monsoon in Cochin port channel is caused by river discharge and tidal inflow respectively</li></ul>		
Kaggle	<b>MNIST Fashion Image Classification of 10 classes with Convolution Neural Network with Keras framework (Deep Learning)</b>		Nov' 19
Overview	<ul style="list-style-type: none"><li>Dataset is collected from <b>MNIST</b> of Zalando's article images—consisting of a training set of <b>60,000</b> examples and a test set of <b>10,000</b> examples. Each example is a <b>28x28</b> grayscale image, associated with a label from <b>10 classes of different fashion apparels</b>.</li><li>Keras with tensorflow as backend is used for building neural network models.</li><li>Neural Network of 2 layer with <b>RELU</b> and <b>SOFTMAX</b> as activation function is developed.</li><li><b>One hot encoding</b> is done as data preprocessing.</li><li><b>Convolution Neural Network</b> model is developed of different layers and with and without <b>padding</b>.</li><li><b>RELU</b> and <b>SOFTMAX</b> is used as activation functions.</li><li>Optimizer used is <b>ADAM</b> and loss used is <b>Categorical Cross Entropy</b>.</li><li>Models are trained with different batch size and epochs to reduce overfitting and time consumed.</li></ul>		
Achievements	<ul style="list-style-type: none"><li><b>Accuracy</b> is used for model evaluation.</li><li>Neural Network with 2 layers has test loss of <b>0.317</b> and accuracy of <b>88.78 %</b></li><li>CNN with 3 layers of convolution and without padding has test loss of <b>.21</b> and accuracy of <b>92.54 %</b></li><li>CNN with 5 layers of convolution and with padding has test loss of <b>.177</b> and an accuracy of <b>93.79 %</b>.</li></ul>		

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