# DATA SCIENTIST INTERN
Professional Summary
Skills

- Python, C++, Java, C#, Pytorch, Tensorflow, Keras, SQL, Flask, Hadoop, AWS.
- Tools: OpenCV, Pandas, Numpy, Sklearn, NLTK, Gensim, Matplotlib, Seaborn, Scipy.
- Artificial Intelligence
- C++
- Clustering
- Chi
- Credit
- Database
- Edge
- Features
- Forecasting
- Image Processing

- Image
- Java
- C#
- Exchange
- Natural Language Processing
- NLP
- Network
- Network File System
- Oil
- Predict
- Profit
- Proposal
- Python
- Real Estate
- Risk Analysis
- SQL
- Statistics
- Valuation
- Vision

Work History
Data Scientist Intern , 06/2020 to 08/2020
Ascend Learning â€" Orlando , FL

- Reduced the Median Average Percentage Error (MAPE) of the Real Estate Valuation Price model by 0.8% to 6.54% using a regression model, including the features engineered from images with the existing model.
- Adopted Room Scene Recognition from a low-resolution network (Resnet 50) to a high-resolution object detector (Faster RCNN with Inception Resnet V2 as the base network) with a F1 Score of 83.4%.
- Automatically flagged blurred, under-exposed and over-exposed images when customers upload the images of their properties by variance of laplacian and brightness measure using the histogram of pixels.

Data Scientist , 06/2017 to 07/2019
Yahoo! â€" Lockport

- Real Estate Property Disposition (Regression, Decision Trees, Clustering, Risk Analysis).
- To increase profit and minimize risk during property disposition, built a predictive model to compute Haircut percentage of a RE property with a RMSE of 3.86%, with Light GBM model.
- Researched and implemented a regression exchange algorithm, for partitioning property disposition dataset into 3 clusters, yielding an R-squared value of 0.821.
- Real Estate Valuation model from Images (Deep Learning, Object Detection, Image Processing, OpenCV).
- Developed pipelines to analyze and extract features from images to estimate the luxury level of houses in USA.
- Researched and designed the architecture of the SSD object detection model to dynamically add new classes to the model preventing Catastrophic Forgetting.
- Reduced the total training time to 1/12th of the time taken to train all the classes from scratch.
- Cut down the manual efforts for labelling to 5% by means of Selective Search region proposal algorithm, K-Means clustering to get bounding boxes on images and group them to be used as labels for object detection algorithms.
- Implemented a system to locate the vanishing points of a room in an image with the help of Canny Edge Detector and Hough Transform, helped in correcting the aspect ratio of the objects detected.

Data Scientist Intern , 01/2017 to 05/2017
Ascend Learning â€" Philadelphia

- To find the location of a house, with the help of the seller's comments of that house, built a NLP model using bag of words and TF-IDF (Term Frequency-Inverse Document Frequency) as the features.
- To predict if the property was over-valuated, built a LSTM (many to one) model by using the vector of sentiment score of all the words in the description of a property as input features.

12/2019

- Decline Curve Analysis (Python, Forecasting, Cost Deduction, Ensembled trees) Jun 2019.
- Analyzed and forecasted the oil and gas quantities in over 3000 wells using mathematical models.
- Modeled and forecasted with Ensembled trees, exponential, hyperbolic and harmonic decline functions.
- Used RMSE as loss function, AIC and BIC as a Criterion for model selection.

- Effect of lockdown on Covid-19 cases and deaths (Inferential Statistics, Python, P-Value) May 2020.
- Performed right tailed paired t-test to check if the number of covid-19 cases and deaths decreased in various states in USA after lockdown was imposed.
- Credit Card Fraud Transaction Detection (Machine Learning, Pandas, OverSampling, Decision Trees, Identified fraud transactions with a F1 score of 0.845 with Gradient Boosted trees from imbalanced dataset.
- Used SMOTE for oversampling and Pearson's correlation coefficient, Chi-square test for feature selection.
- Action Recognition Using CNN and LSTM (Computer Vision, NLP, TensorFlow) Nov 2019.
- Detected the action performed in videos by capturing spatial information in frames using Resnet 101 and temporal information between frames using LSTMs with a threefold accuracy of 89.6%.
- Named Entity Recognition (NLP, BERT, Transformer, Python, Pytorch ) Mar 2020.
- Pretrained a BERT model with the Request For Comments text for Network File System implementation.
- Finetuned and built a NER model for identifying Variables, Values and Operations with a F1 score of 0.6125.

Education
Masters : Computer Science , Expected in 05/2021
State University of New York - City , State
GPA: 3.62/4
B.Tech : Computer Science , 05/2017
Amrita School of Engineering - City
GPA: 8.84/10
Work History
12/2019

- Decline Curve Analysis (Python, Forecasting, Cost Deduction, Ensembled trees) Jun 2019.
- Analyzed and forecasted the oil and gas quantities in over 3000 wells using mathematical models.
- Modeled and forecasted with Ensembled trees, exponential, hyperbolic and harmonic decline functions.
- Used RMSE as loss function, AIC and BIC as a Criterion for model selection.
- Effect of lockdown on Covid-19 cases and deaths (Inferential Statistics, Python, P-Value) May 2020.
- Performed right tailed paired t-test to check if the number of covid-19 cases and deaths decreased in various states in USA after lockdown was imposed.
- Credit Card Fraud Transaction Detection (Machine Learning, Pandas, OverSampling, Decision Trees, Identified fraud transactions with a F1 score of 0.845 with Gradient Boosted trees from imbalanced dataset.
- Used SMOTE for oversampling and Pearson's correlation coefficient, Chi-square test for feature selection.
- Action Recognition Using CNN and LSTM (Computer Vision, NLP, TensorFlow) Nov 2019.
- Detected the action performed in videos by capturing spatial information in frames using Resnet 101 and temporal information between frames using LSTMs with a threefold accuracy of 89.6%.
- Named Entity Recognition (NLP, BERT, Transformer, Python, Pytorch ) Mar 2020.
- Pretrained a BERT model with the Request For Comments text for Network File System implementation.
- Finetuned and built a NER model for identifying Variables, Values and Operations with a F1 score of 0.6125.

Data Scientist Intern , 06/2020 to 08/2020
Bickford Senior Living â€" Crawfordsville , IN

- Reduced the Median Average Percentage Error (MAPE) of the Real Estate Valuation Price model by 0.8% to 6.54% using a regression model, including the features engineered from images with the existing model.
- Adopted Room Scene Recognition from a low-resolution network (Resnet 50) to a high-resolution object detector (Faster RCNN with Inception Resnet V2 as the base network) with a F1 Score of 83.4%.
- Automatically flagged blurred, under-exposed and over-exposed images when customers upload the images of their properties by variance of laplacian and brightness measure using the histogram of pixels.

Data Scientist , 06/2017 to 07/2019
Company Name â€" City

- Real Estate Property Disposition (Regression, Decision Trees, Clustering, Risk Analysis).
- To increase profit and minimize risk during property disposition, built a predictive model to compute Haircut percentage of a RE property with a RMSE of 3.86%, with Light GBM model.
- Researched and implemented a regression exchange algorithm, for partitioning property disposition dataset into 3 clusters, yielding an R-squared value of 0.821.
- Real Estate Valuation model from Images (Deep Learning, Object Detection, Image Processing, OpenCV).
- Developed pipelines to analyze and extract features from images to estimate the luxury level of houses in USA.
- Researched and designed the architecture of the SSD object detection model to dynamically add new classes to the model preventing Catastrophic Forgetting.
- Reduced the total training time to 1/12th of the time taken to train all the classes from scratch.
- Cut down the manual efforts for labelling to 5% by means of Selective Search region proposal algorithm, K-Means clustering to get bounding boxes on images and group them to be used as labels for object detection algorithms.
- Implemented a system to locate the vanishing points of a room in an image with the help of Canny Edge Detector and Hough Transform,

helped in correcting the aspect ratio of the objects detected.

Data Scientist Intern , 01/2017 to 05/2017
Company Name â€" City

- To find the location of a house, with the help of the seller's comments of that house, built a NLP model using bag of words and TF-IDF (Term Frequency-Inverse Document Frequency) as the features.
- To predict if the property was over-valuated, built a LSTM (many to one) model by using the vector of sentiment score of all the words in the description of a property as input features.

Skills
- Python, C++, Java, C#, Pytorch, Tensorflow, Keras, SQL, Flask, Hadoop, AWS.
- Tools: OpenCV, Pandas, Numpy, Sklearn, NLTK, Gensim, Matplotlib, Seaborn, Scipy.
- Artificial Intelligence, C++, Clustering, Chi, Credit, Database, Edge, features, Forecasting, Image Processing, image, Java, C#, exchange, Natural Language Processing, NLP, network, Network File System, oil, predict, profit, proposal, Python, Real Estate, Risk Analysis, SQL, Statistics, Valuation, Vision