The Journey of Entry Level Data Scientists …

By Arun Prakash, Garima Jena, Haneesh Reddy, Sneha Komma.

The world of marching science and galloping technology demands tough decisions in life to stay special among the whole lot. That one special decision in life of aspiring the role of a data scientist, made individually, enabled us to come together and take up CBA, an Executive Programme in Business Analytics from Indian School of Business, to learn, share, challenge, wonder, aspire, investigate, question, enjoy the various facets of Data Science. The complete-diet to ensure all the nutrition for a healthy Data Scientist prescribed by most eminent personalities in the field have given us the wonderful opportunity of sharing the joy of learning and experimenting. We are presented with an opportunity to showcase our classroom learning in a real life situation. It's imperative to apply the theories and concepts taught, learnt in the class to a real environment and this project enabled us to do so.

We as a team participated in the open, world level data competition, Kaggle, for a project titled "Facial Key point Detection". The main purpose of the project is to learn how to analyze and predict key point positions on face images, which can be used as a building block in several applications such as tracking faces in images and videos, analyzing facial expressions, detecting dysmorphic facial signs for medical diagnosis, domestic biometrics/face detection, terrorist detection activities and the like. The objective of the task is to predict 15 key point positions on face images using a training data set of 7043 images along with X, Y coordinates of 15 facial key points (for example, left eye center, nose tip, etc). The unsupervised machine learning model gets trained on the training data set and enables us to predict the facial key points on the test data set consisting of 1742 images. The model uses a K-Means clustering along with different techniques like covariance scores on Z normalized patches of images to figure out the right key points. The model is finalized after having experimented with different clustering techniques, Gaussian Mixtures of models, the use of different patch scales of images like 11*11, 15*15, 21*21 for cluster formation and comparison, different scoring techniques Euclidean distance, Covariance, Correlation etc.

Using the model built we are able to predict all the facial key points challenged by Kaggle competition. The Root-Mean-Square Deviation (RMSD) of the predicted pixel coordinates from that of actual is around 3 pixels and we are happy to get listed in the top 4 ranks in the leader board. The project also bags the "Best Project Award" at ISB, as part of the project done for the Big Data Analytics (Data Management II) course. The chosen task is tricky, quite real time and is ever evolving. Detecting facial key points is a challenging problem and we initially had some difficulty how to go about it. With the team, constantly enlightened by the Professor, coupled with hard work and curiosity to learn, paved way to success. Understanding the problem, deciding the algorithm, brainstorming and experimenting with different methodologies, reaching every decided milestone added more spice to our journey. The overall journey is indeed cherishable and thoroughly enjoyed! We are in the process of implementing Map-Reduce to improve the performance of the algorithm.

We are grateful to ISB for letting us gain this fruitful experience and the plethora of learning. Our special thanks to Dr Shailesh Kumar for the patient guidance, encouragement and advice he has provided throughout the project. He is one big inspiration for us and all these achievements would not have been possible without his mentorship. Special thanks to Dr. Yoshua Bengio of the University of Montreal who provided us the data set for this competition.

Our team is a right mix of talents - a Techno-Functional Business Analyst, a Business Enthusiast, a Developer and an Associate Analyst. To know more about please visit our profiles at

Arun Prakash - http://www.linkedin.com/profile/view?id=269623547
Garima Jena - http://www.linkedin.com/profile/view?id=257730803
Haneesh Reddy - http://www.linkedin.com/profile/view?id=113107359
Sneha Komma - http://www.linkedin.com/profile/view?id=178413607