**CSE 519 -- Data Science**

**Quant Shop Video Review Form**

Name: Rohan Vaish Email: rvaish@cs.stonybrook.edu

ID Number 111447435

Video episode title: Winning the Baby Pool Episode number (1-8): 3

Of the eight episodes, how many have you previously watched?  (0-7): 0

Give the month and year you started your graduate studies at Stony Brook: August 2017

Did you learn anything by watching it?  (Y/N): Y  Describe stuff you did learn below (if Y):

I learnt a whole lot about Data gathering and analysis watching my first episode which predicts the baby weight just after birth. Getting or developing the right dataset for our model to work on, is the most tedious and important task. Different data sets have different parameters (for example smoking factor was not present in California dataset but was present in the New York one) which makes it difficult to develop the baseline parameters for our model. Also, error visualization(distribution) on parameters one at a time helps us in the big picture, which makes our prediction mechanism less complex.

What did you learn from looking at their project report?

They used ethnicity, sex of the child, age group and location as the baseline model parameters but later went on to apply Sammy’s parameters to Random Forest Regressor (with the least root mean square error and highest R2 score). In the end they predicted the child weight at 7 lbs and 10 ozs and delivery date as December 8th. Even while the delivery date came out to be December 17th, the baby weighed 7 lbs 3 ozs which was very close to their prediction. Also, I learnt that previous/current disease in the mother, smoking habits, drinking habits, sex of the child, race of the parents, prenatal care can also affect the baby weight, which they included in their models in the later stage giving this near to accurate prediction. Studying root mean square error while passing different predictions to our model seems to be as one of the most important tasks to study the quality of our model vs the other models.

Rate program quality from 1 (terrible) to 10 (amazing):8 Was it worth watching (Y/N)? Y

What could we do to make the videos better, and more educational?

1. More visual depiction of the data sets.
2. More insight about how to choose the right data sets and models.
3. Incorporating the project reports in the videos in short

|  |
| --- |
|  |