Model to be created: Y = aX1 + bX2 + cX3 +dX4 + eX5 + fX6

Dependent variable / Response variable / Predicted variable – y variable

Independent variables / Predictor Variables / X variables / Features / Dimensions / X – X1 , X2, X3, X4 based on which the prediction has to be done

Regression problems: Prediction problem where y is continuous

Classification problems: Prediction problem where y is categorical (categories)

Predictive Analytics is also called as Supvervised Learning – Means model is learning through the training data. Model is learning the parameters of the model: y = aX1 + bX2 + c. The model is learning the parameters of the model a, b and c. Once we have trained the model, or obtained the parameters of the model (which include the coefficients (or weights) + the bias (or intercept)), we can predict the values of y for any new entry.

Natural Language Processing: deciphering the natural language – text / speech based analysis. Eg. google home, amazon alexa, Cortana, chatbots, apple siri, spam / ham classification of emails, sentiment analysis

Big Data – volume, velocity of data, variety

Volume – large datasets

Velocity – the data is coming very quickly and needs to be analysed in the real time

Variety – strucutured, unstructured: speech, text, pictures / images (google lens, captcha, vehicle number plate at the gate security, facial recognition, facebook person tagging, instagram. Each image is an array of pixels 8 X 8 X 3 channels (RGB) – 1 92 features.) Video – comprises of multiple frames / images)