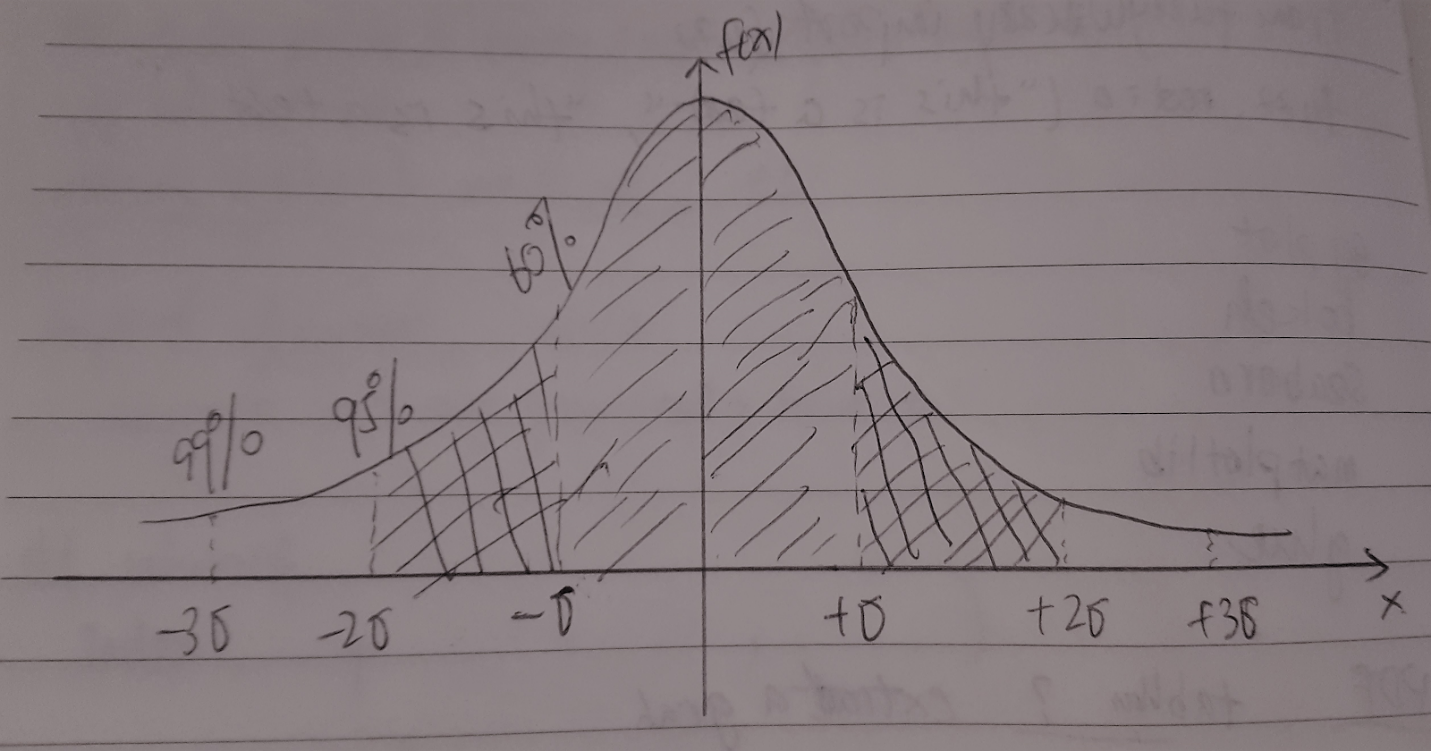
Dow Jones Index proposal

1.    What is the problem you want to solve?

The problem to solve is to analyze the historical data of Dow Jones Index Average(DJIA) and predict the close value of DJIX. The features used in the model includes the day of the week, the month, the quarter and the year; the open, close, high and low value of the previous days;

2.    Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

Stock investors care about this problem as it might affect their gain and loss in their investment. The clients will decide to buy or sell DJIA stocks based on the prediction of the tomorrow value is higher or lower than today’s value. The accuracy is more than 50% for the direction of up and down. The 99% of the confidence interval should be above than 50%, as the figure shown below.



3.    What data are you going to use for this? How will you acquire this data?

The data of this project is the 10-year historical data of Dow Jones Index. It can be acquired from yahoo finance page.

4.    In brief, outline your approach to solving this problem (knowing that this might change later).

The approach to solve this problem is shown as below:

1. Analyze the data pattern;
2. Load the data;
3. Clean the data by
4. Plot the data over time;
5. Exploratory data analysis, which includes statistics as mean,standard variation and etc. by different category(features);
6. Create model SGDRegressor with similar codes as below:

sgd = SGDRegressor()

sgd.fit(X, y)

X = features dataframe

y = DJIA close that you want to predict

Cross validation

using y\_predicted = sgd.predict(X)

y\_predict - y

5.    What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

Deliverables includes code and a slide deck,