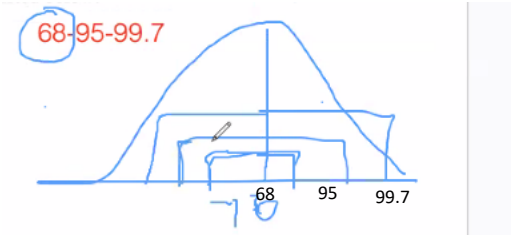
Distribution

Thursday, February 26, 2026 11:37 AM

Normal or Gaussian Distribution – Normal data like scores

Probability Distribution Function : possible outcomes can be calculated

Rule : how data can be identified – Text boc(68-95-99.7)



Remaing people 40 are failure and 2-3 percent outliers

Types of the probability distribution

There are many different types of probability distribution. Some of them which we will be covering in this blog are listed below:

- Normal Distribution Or Gausian
- Bernoulli's Distribution Give a single output value and experiment is only performed once -winning a toss
- Binomial Distribution N*Bernoulli's is binomial
- Uniform Distribution =Distribution has all your values have equal probability
- Poisson Distribution = No of times an event has occered in a particular interval

CLASSIFICATION MODEL

Correlation between the columns need to be handled with heat map which will impact the data
Overfitting , Underfitting

Feature engineering : creating new input features from existing data to improve model performance.
Feature selection : choosing the most important features and removing unnecessary ones.

Crossvalidation- split dataset into training dataset and test dataset

Categorical to numerical conversion

```
# method to convert categorical data to numerical data
def categorical_to_numeric(x):
    if 0 <=  x < 6:
        return 0
    elif 6 <= x < 13:
        return 1
    elif 13 <= x < 19:
        return 2
    elif 19 <= x < 24:
        return 3
```