

# PREDICTING BREAST CANCER

## Fine Needle Aspiration

BY: OMEGA MARKOS



# BREAST CANCER STATISTICS

What is the risk of being diagnosed with breast cancer?

**13%**

How many women alive today ?

**3.8 million**

How many cases are expected to occur in 2019?

**268,600** -women

**2,670** - men.

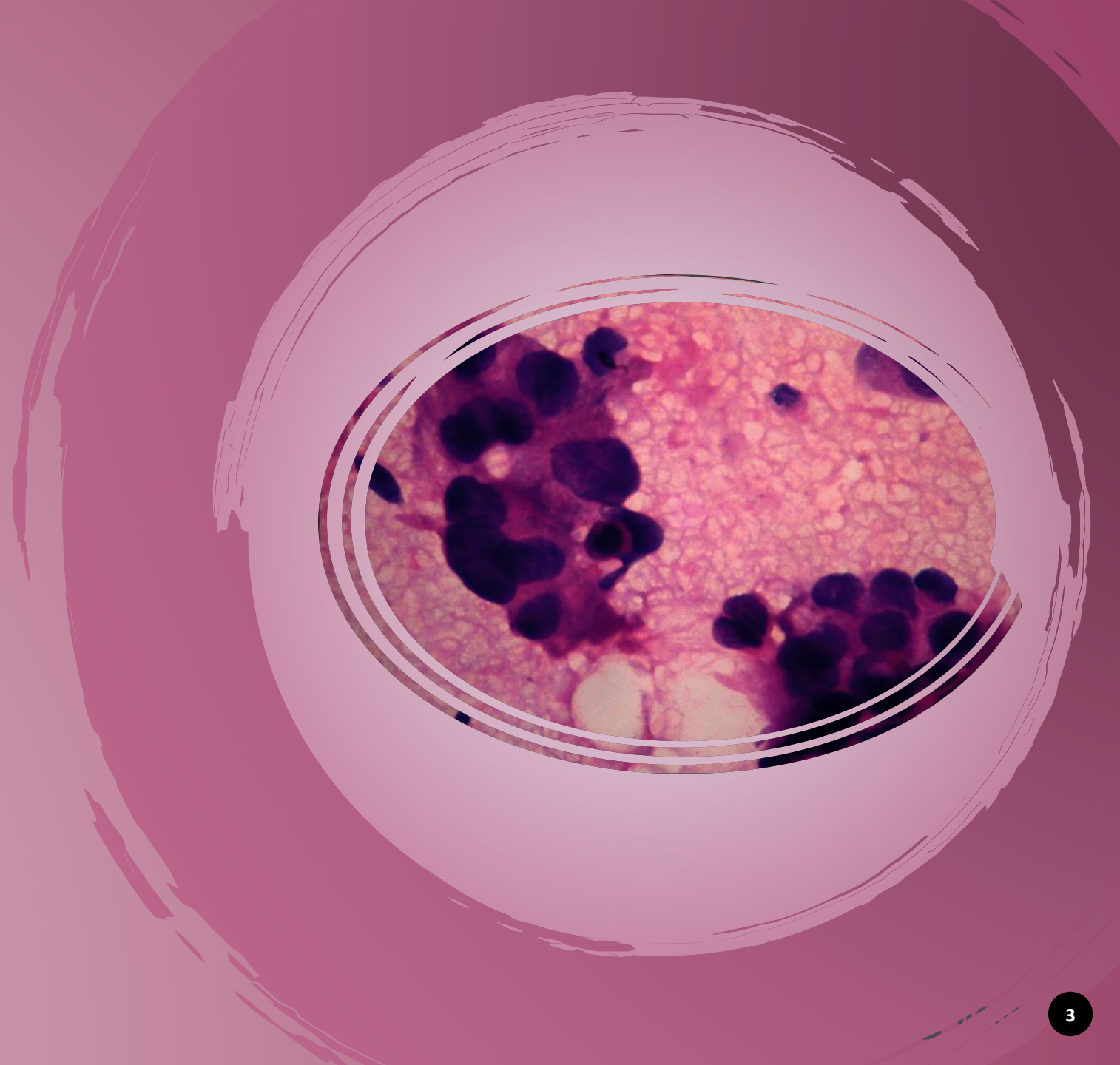
The incidence rate for invasive breast cancer has been slowly **increasing** since 2004.



# THE PROBLEM

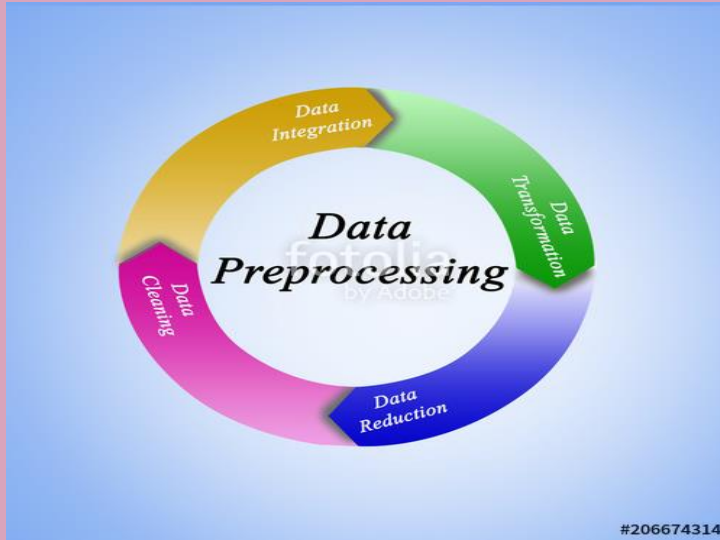
## Inconsistent diagnosis

- Inexperienced cytologist
- Overlap of cytological features of benign vs. Malignant



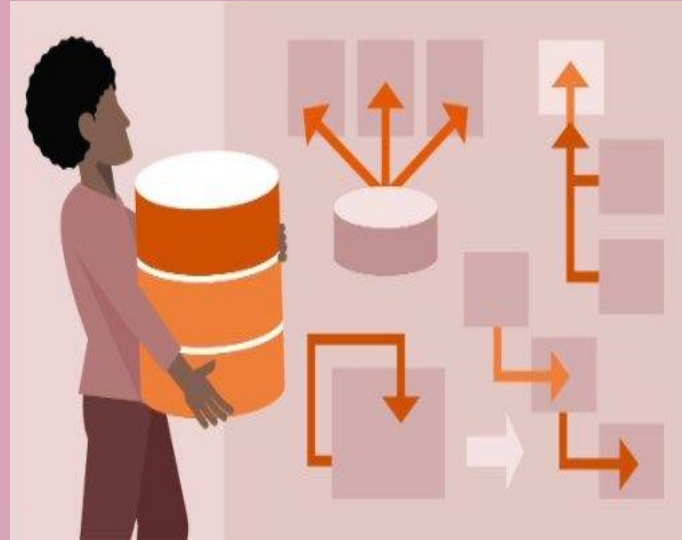
# Methodology

Three step methodology



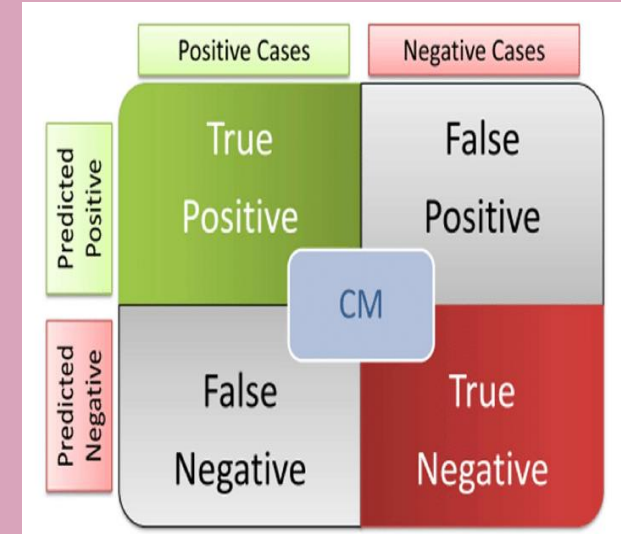
## Data Preprocessing

Data cleaning  
Feature selection



## Data Modeling

Fit different models



## Model Evaluation

Evaluate the models  
for accuracy



# Correlated features

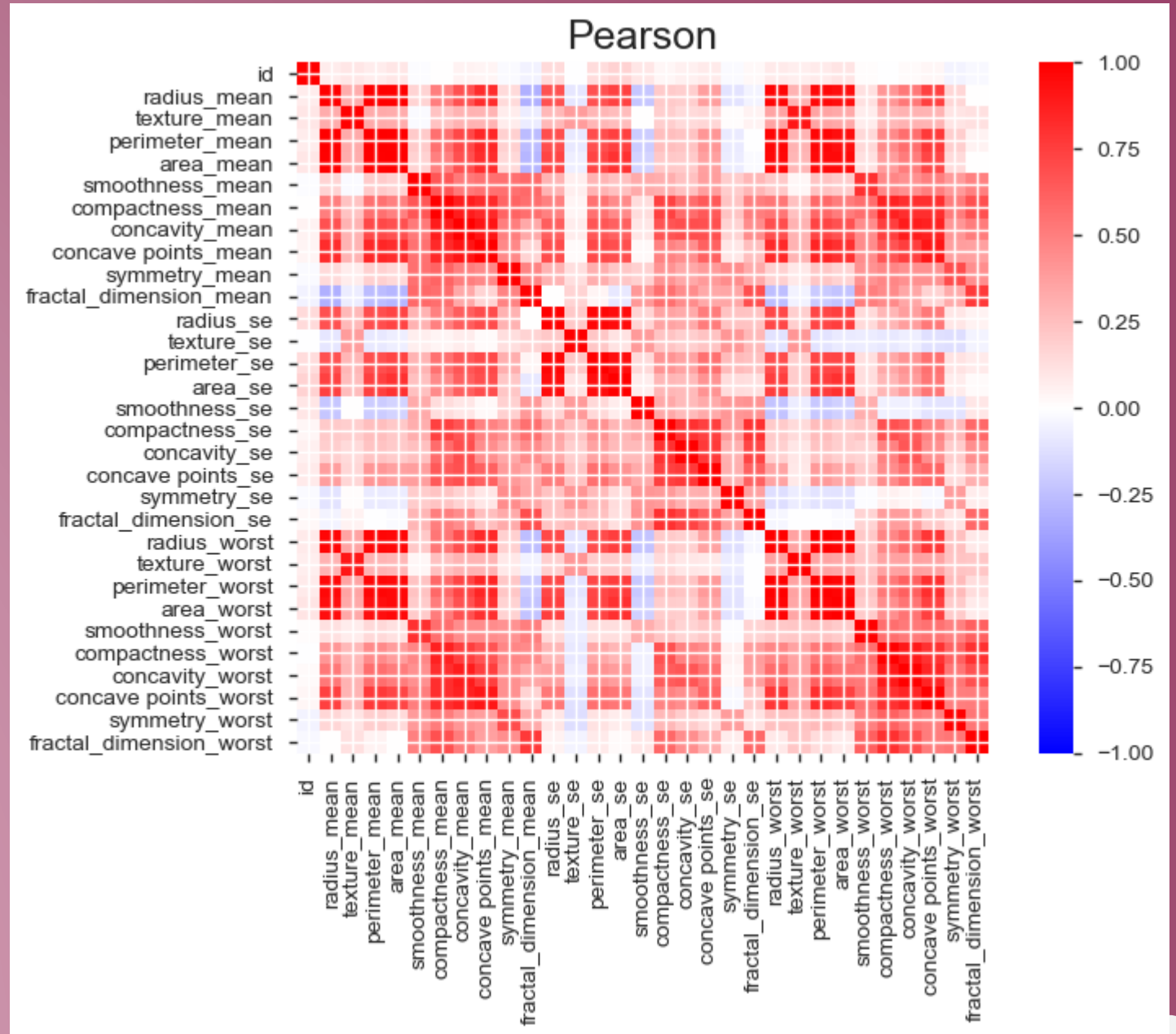
1.radius\_mean, perimeter\_mean and area\_mean.

2.concavity\_mean, concave points\_mean & concave points\_worst .

3.radius\_worst ,perimeter\_worst and area\_worst

4.radius\_se, 'perimeter\_se and area\_se

5.texture\_mean & texture\_worst



# IMPORTANT FEATURES

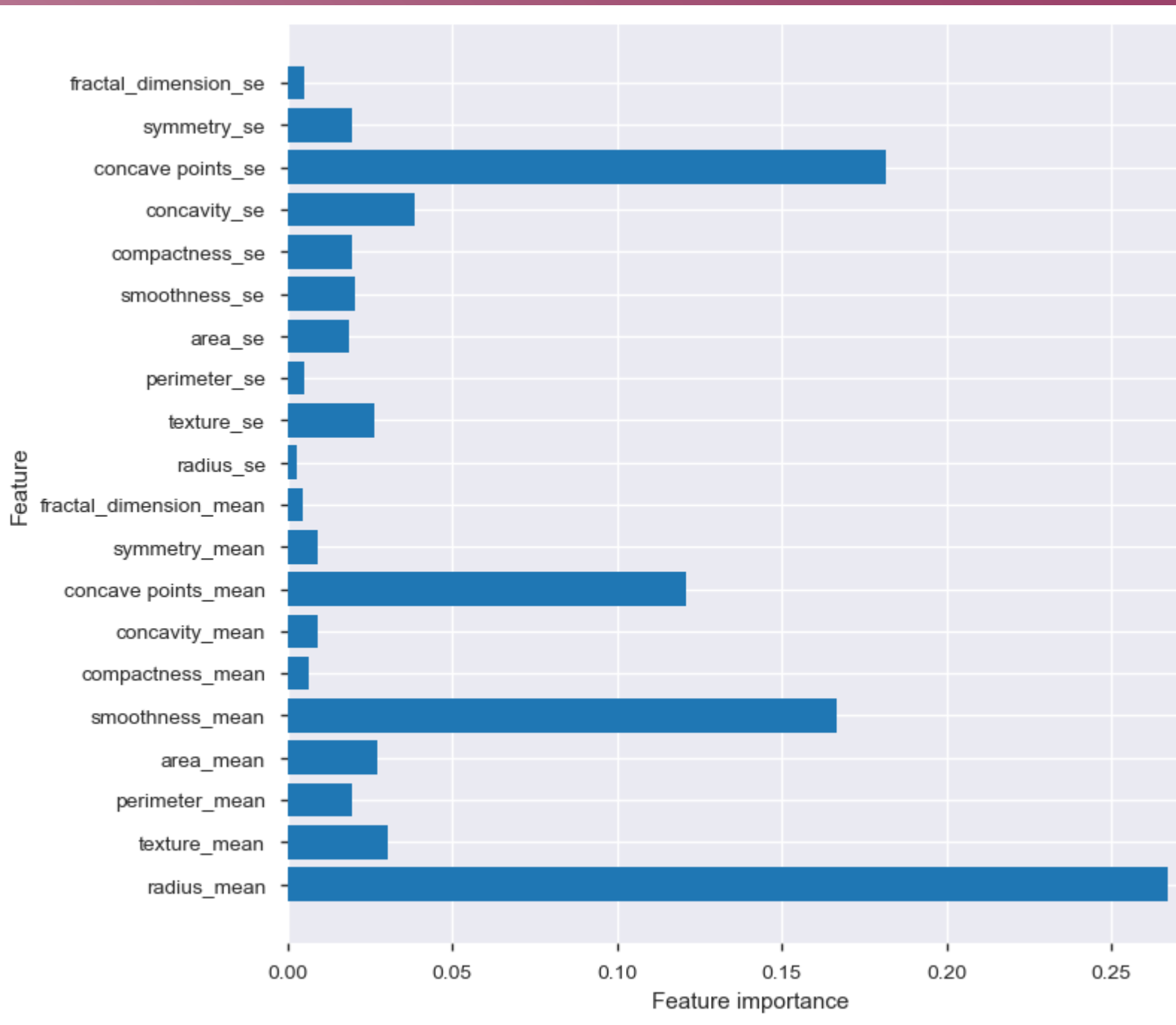
**Radius\_mean**

**Concavity\_point\_se**

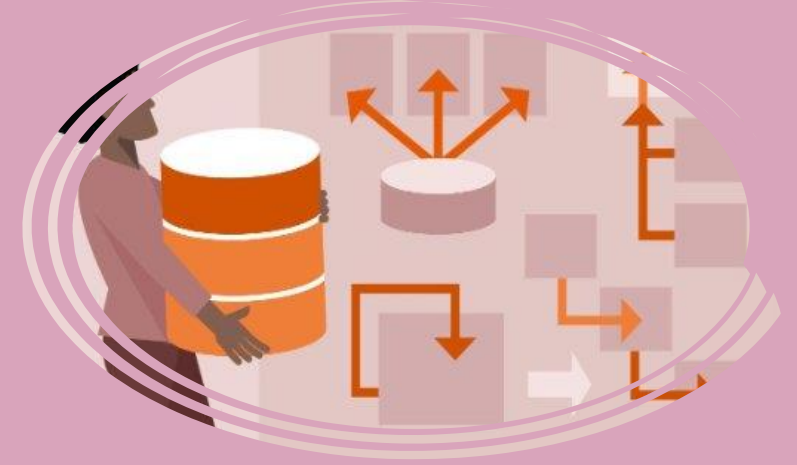
**Smoothness\_mean**

**Concavity\_point\_mean**

**Concavity\_se**



# Model Accuracy



Model Name	Model Accuracy %	Recall Score %
Logistic Regression	97	98
Random Forest	96	98
Adaboost	100	93
Naïve Bayes	91	90

Test best classifier -Logistic regression

# Recommendation

## 1.The 20 features:

- |                          |                            |
|--------------------------|----------------------------|
| 1.radius_mean            | 11.compactness_se          |
| 2. texture_mean,         | 12.concavity_se            |
| 3.smoothness_mean,       | 13.concave points_se       |
| 4.compactness_mean,      | 14.symmetry_se             |
| 5.concavity_mean,        | 15.fractal_dimension_se    |
| 6.symmetry_mean,         | 16.smoothness_worst,       |
| 7.fractal_dimension_mean | 17.compactness_worst       |
| 8. radius_se             | 18.concavity_worst         |
| 9.texture_se,            | 19.symmetry_worst          |
| 10.smoothness_se         | 20.fractal_dimension_worst |

2.Logistic regression the best classifier with 97% accuracy and 98% recall.

## Future work

More detailed descriptive analysis on the important & correlated features





# THANK YOU

OMEGAGHIRMAY@GMAIL.COM ✉