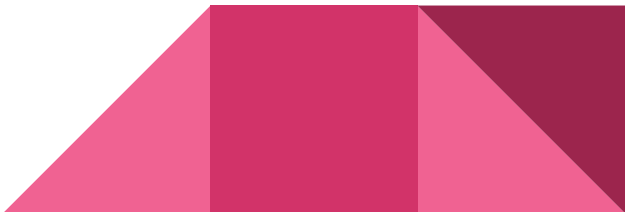


# Fake News Detection Using ML

Yash Mirge (172010021)

# Topics

- Introduction
  - Existing system
  - How to analyze a news ?
  - Features
  - Consequences
  - Types of Fake news detection
  - Technique used by detecting fake news
  - References
- 

# Introduction

- Stories that look like real news stories but are propaganda and disinformation.
- Fake news and lack of trust in the media are growing problem in our society.
- Fake news typically appears on website that look professional . the stories often relate to topics and people who are trending on google and facebook . the stories usually have outrageous headlines designed to get people to click .



# Existing System

- Most of people has been focusing on classifying online reviews and publicly available social media post .
- Fake news hide important context information .
- Particularly since late 2016 during the American Presidential election , the question of determining 'Fake news' has also been the subject of particular attention within the literature




# How to analyze a news ?

**A** → **Bob** → **A**

**C** → October 15 2017 at 15:00 pm → **C**

**B** → Trump is getting support from every leader, and that's the support that will make him grow great and strong. These elections will bring an immense change in our country.


**B** → 

**B** → Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement

**A** → WWW.DAILYPRESSER.COM | BY THE AMERICAN PATRIOT

**C** → Like Comment Share Embed 125 Top Comments

**D** → 

**D** → 

**A: Creator/Spreader**  
**B: News Content**  
**C: Social Context**  
**D: Target**

# Features Of Fake News Detection

- Language Features
- Lexical Features
- Psycholinguistic Features
- Semantic Features



# Language Features

- Sentence-level Features (Bag-of-words approach , part-of speech approach)
- Include number of words and syllables per sentence as well as tags of word categories  
(Such as noun , verb , adjective)



# Lexical Features

- Meaning of a word or a phrase as it is actually used by people
- Include character and word-level signals, such as unique words and their frequency in the text.
- Also implement some pronouns , verbs , hashtags , all punctuations counts , etc..





# Psycholinguistic Features

- Linguistic Inquiry and word count is a dictionary based text whose output has been explored in many classification tasks , include **Fake News Detection** .

## Semantic Features

- Capture the semantic aspects of a text are useful to infer patterns of meaning from data .



# Consequences

- Fake news sometimes gets more views than real news .
- Sometimes politicians and professionals journalists even quote fake news stories !
- Sometimes people engage in illegal and violent behaviour as a result of believing a fake news story .



# Types of fake news

- Visual - based
- User - based
- Knowledge - based
- Style - based



## Visual - based

- These fake news posts use graphics a lot more in as content , which may include images , videos or combination of both .

## User - based

- This type fake news are generated by fake account and is targeted to specific audience which may represent certain age groups gender,culture,political,affiliations .



# Knowledge - based

- These types of posts give scientific explanation to the some unresolved issues and make users to believe it is authentic . For example natural remedies of increased sugar level in human body .

# Style - based

- This types of posts are written by pseudo-journalists who pretend and copy style of some expert journalists.



# Techniques used by detecting fake news

- Naive bayesian classification
- Random Forest
- Support Vector Machine



# Naive Bayes Classification

- Bayes' theorem with an assumption of independence between predictors.
- In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.
- Naive Bayesian model is easy to build and particularly useful for very large data sets.
- For example, a fruit may be considered to be an apple if it is red, round, and about 3 inches in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier would consider all of these properties to independently contribute to the probability that this fruit is an apple.



Bayes theorem provides a way of calculating posterior probability  $P(c|x)$  from  $P(c)$ ,  $P(x)$  and  $P(x|c)$ . Look at the equation below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Diagram illustrating the components of Bayes' theorem equation:

- $P(c|x)$  is labeled as **Posterior Probability**.
- $P(x|c)$  is labeled as **Likelihood**.
- $P(c)$  is labeled as **Class Prior Probability**.
- $P(x)$  is labeled as **Predictor Prior Probability**.

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \cdots \times P(x_n | c) \times P(c)$$

- $P(c/x)$  is the posterior probability of *class (target)* given *predictor (attribute)*.
- $P(c)$  is the prior probability of *class*.
- $P(x/c)$  is the likelihood which is the probability of *predictor* given *class*.
- $P(x)$  is the prior probability of *predictor*.



# Random Forest

- The **Random Forest** Algorithm is composed of different decision trees, each with the same nodes, but using different data that leads to different leaves.
- It merges the decisions of multiple decision trees in order to find an answer, which represents the average of all these decision trees.
- The random forest algorithm is a **supervised learning** model; it uses labeled data to “learn” how to classify unlabeled data.
- The Random Forest Algorithm is used to solve both regression and classification problems



# Features for fake news representation

## Creator/User-based Features

User profiling features

User credibility features

Behavior-based features

## News Content-based Features

Linguistic & Syntactic features

Style-based features

Visual-based features

## Social Context-based Features

Network-based features

Impact-based features

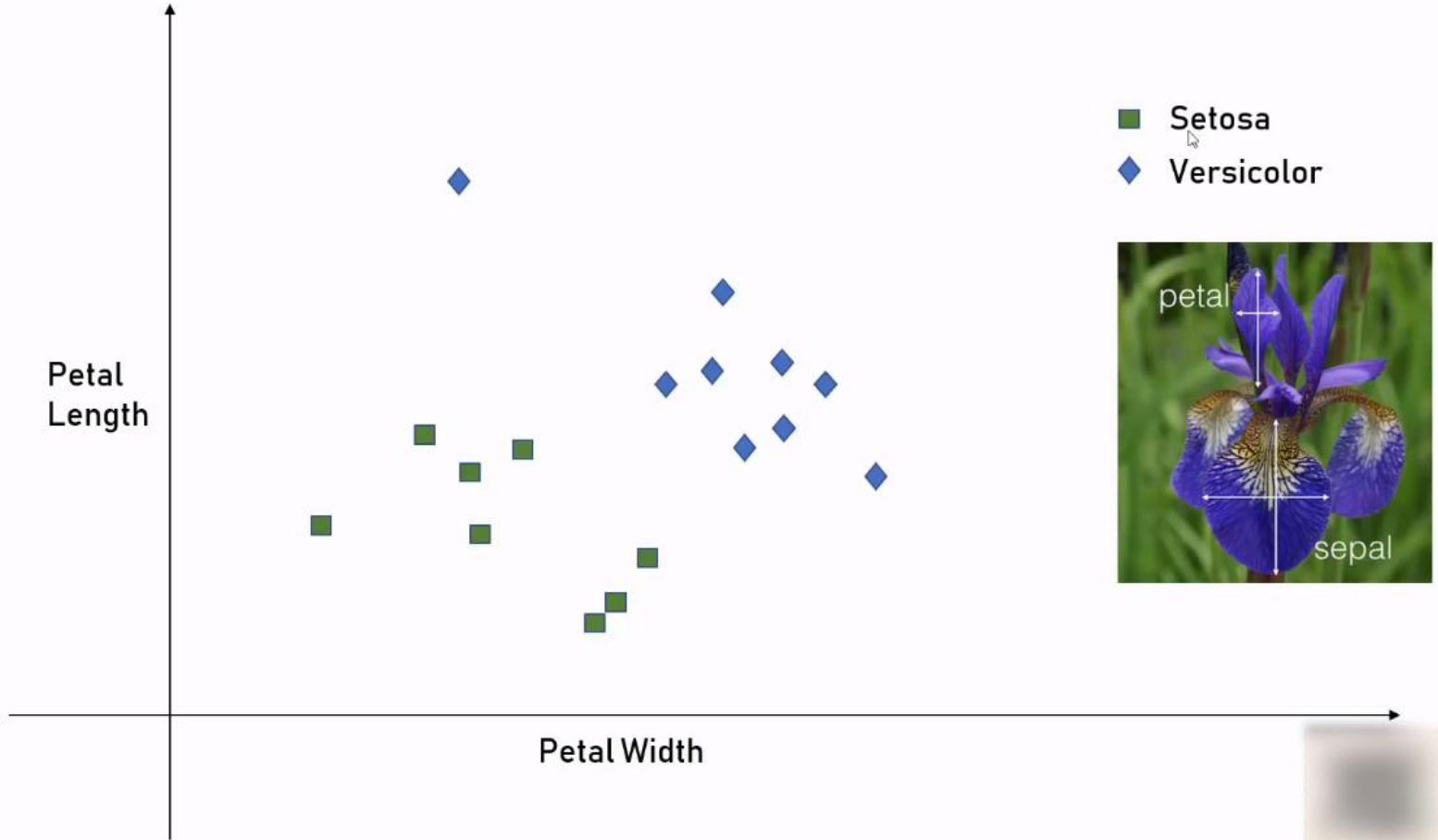
Temporal-based features

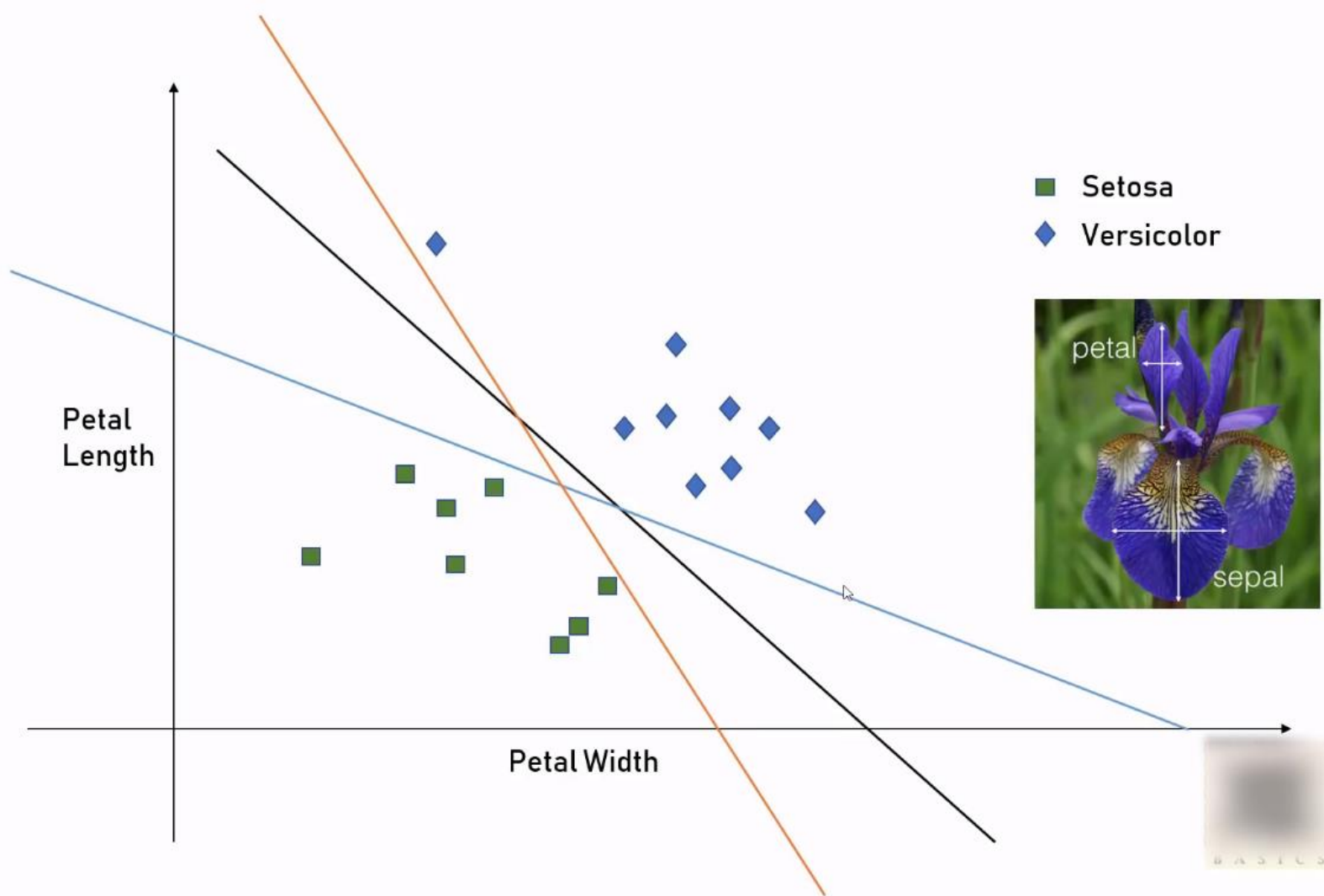
# Support Vector Machine

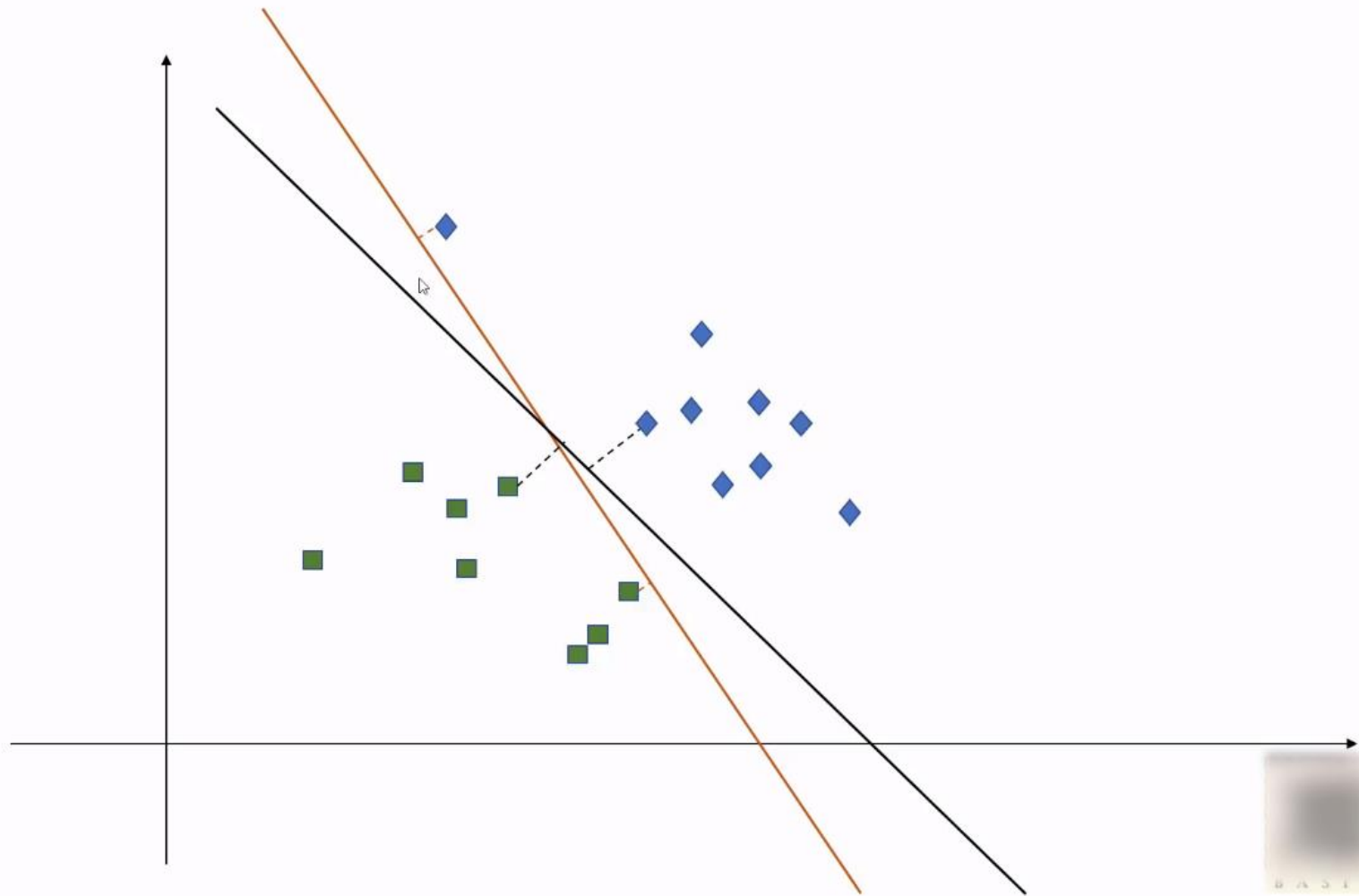
- “Support Vector Machine” (SVM) is a supervised ML algorithm which can be used for both classification or regression challenges
- We perform classification by finding the hyperplane that differentiate the two classes very well
- SVM is robust to outliers
- Maximizing the distances between nearest data point (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as **Margin**.

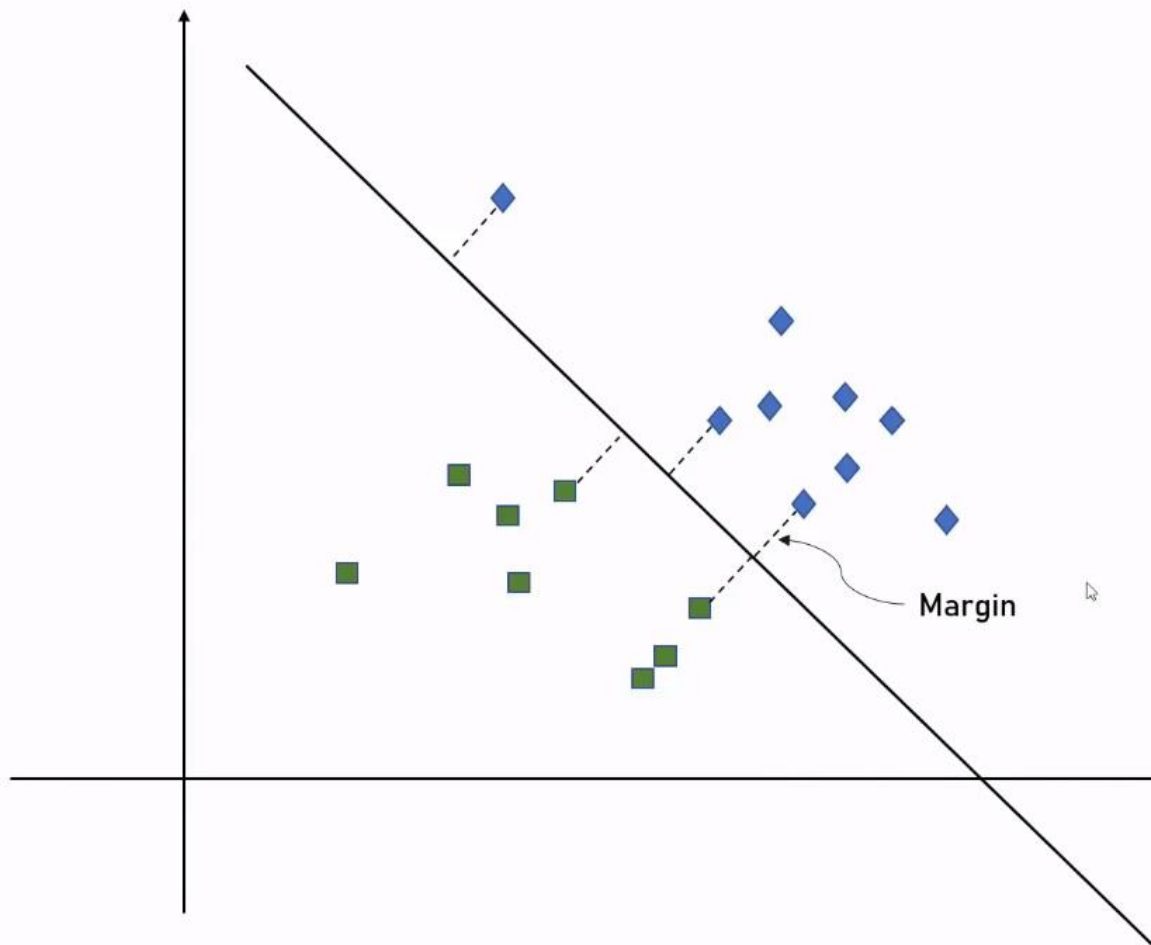


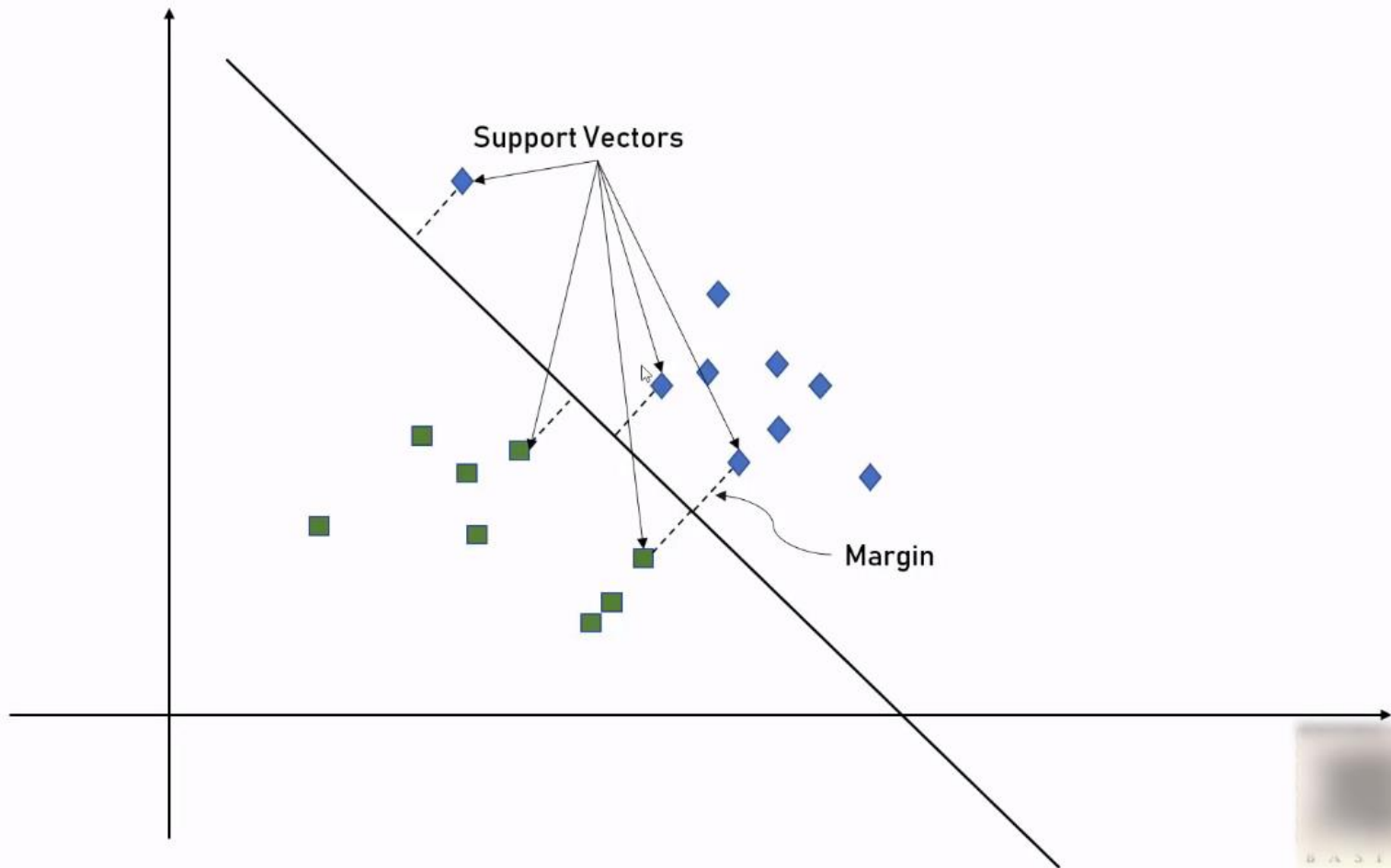
# Finding the Hyperplane..



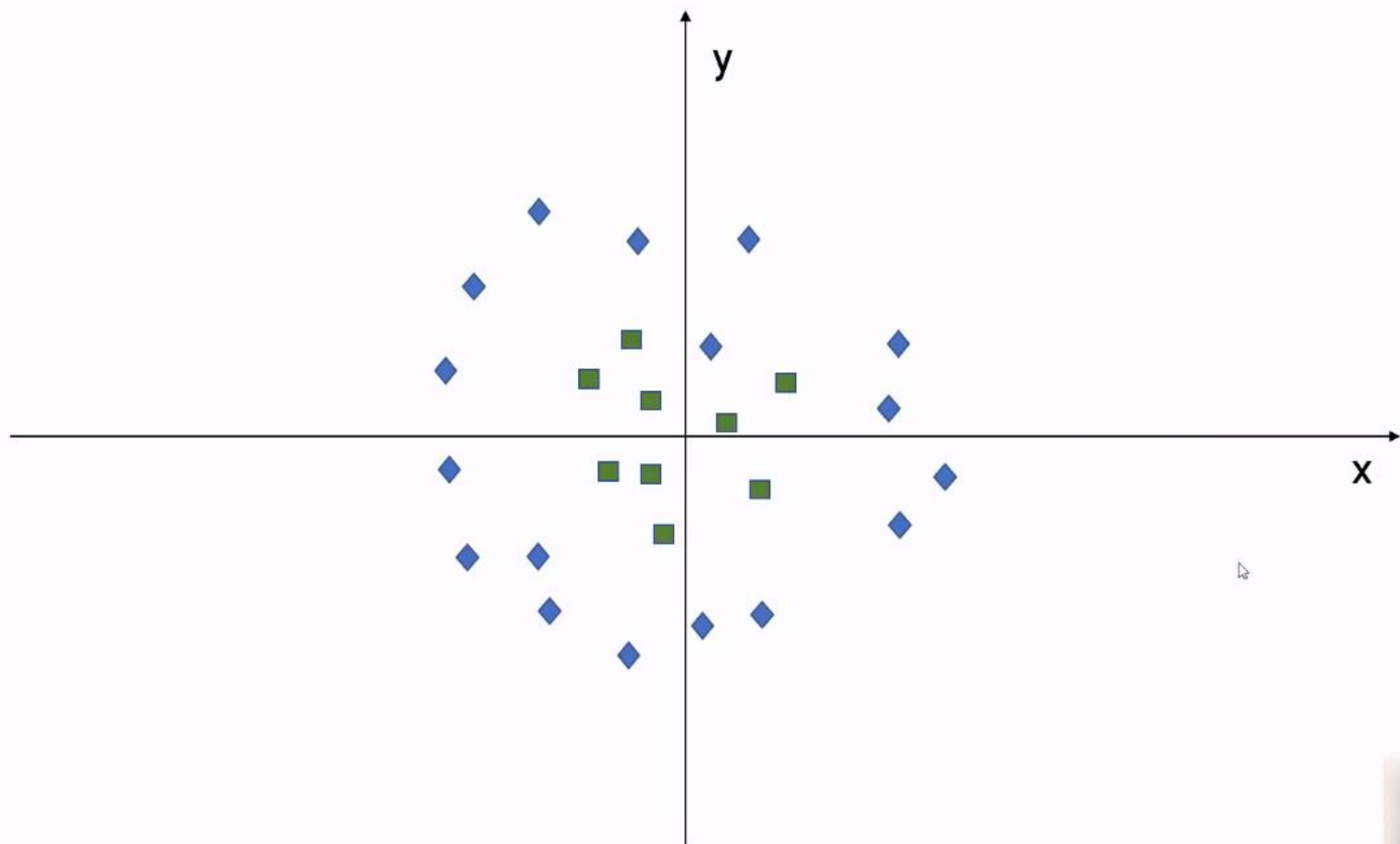




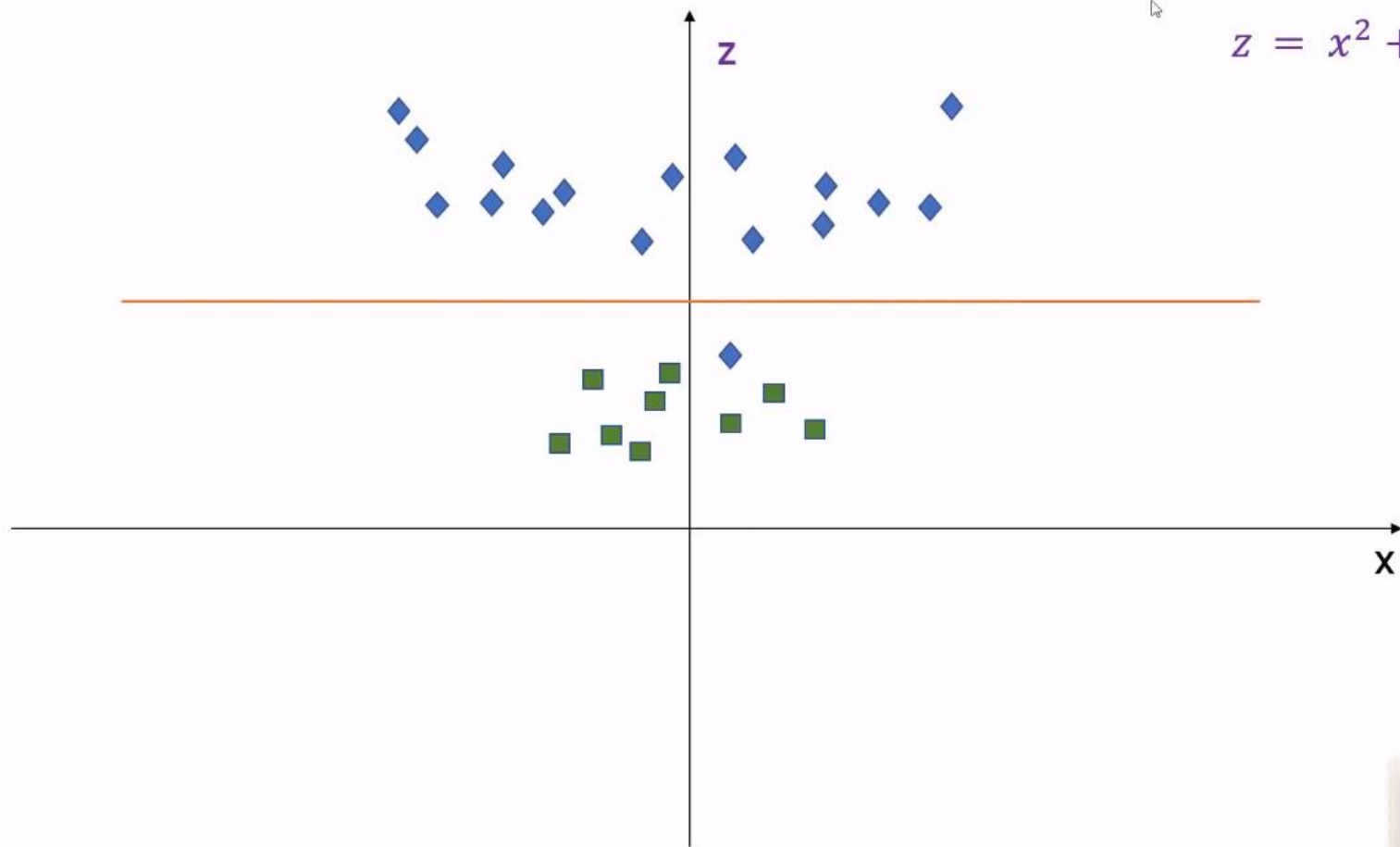


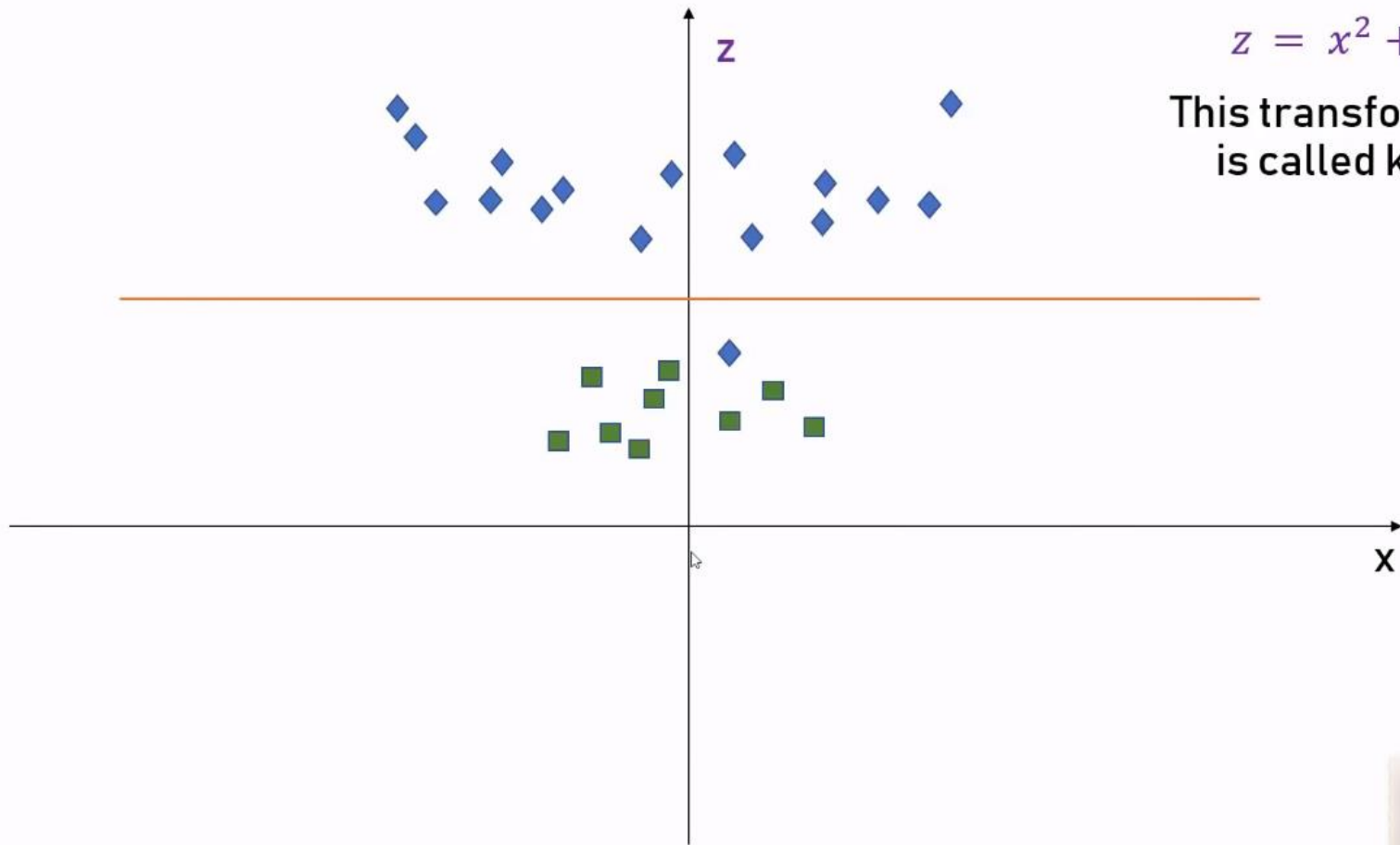






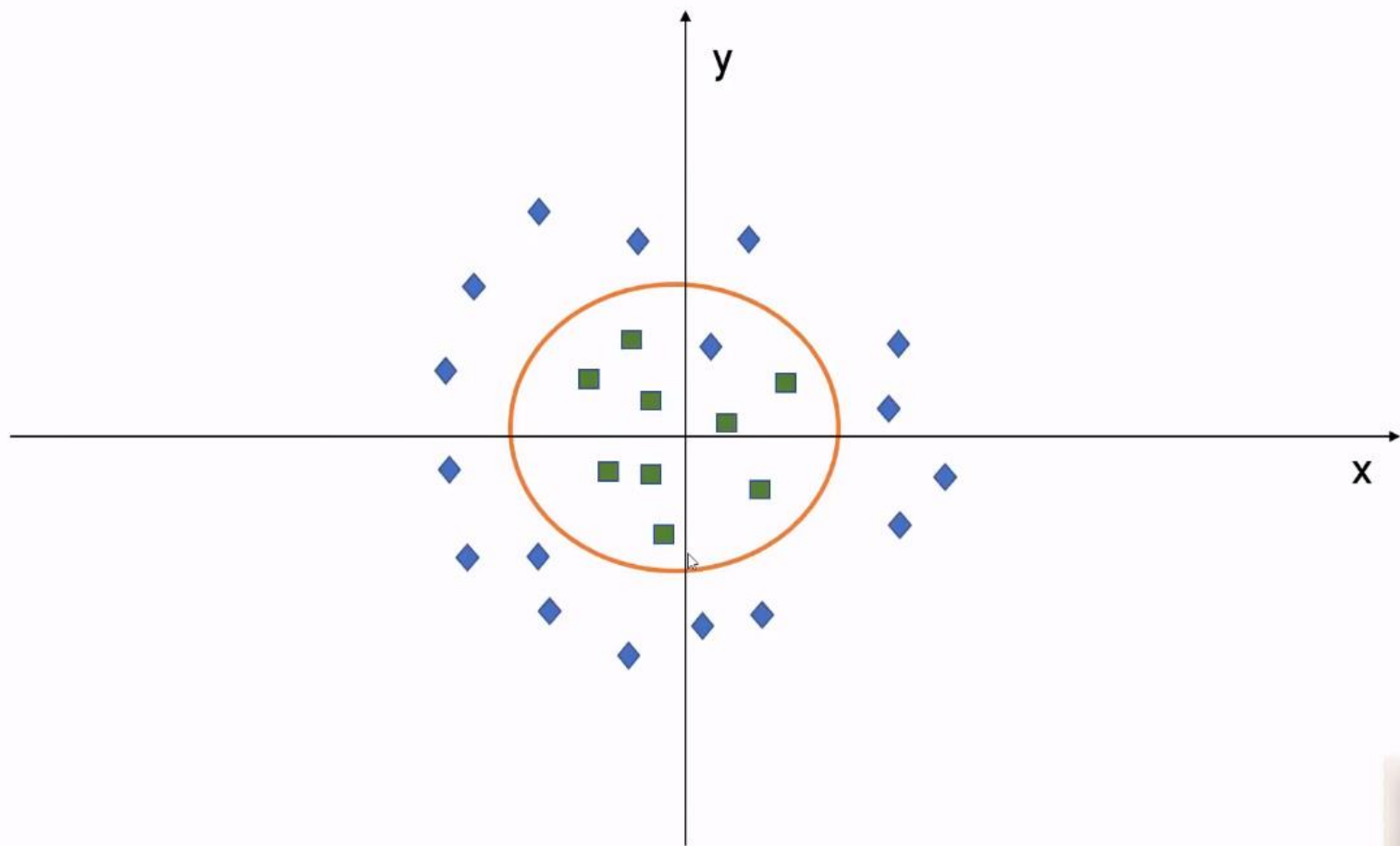
$$z = x^2 + y^2$$





$z = x^2 + y^2$   
This transformation  
is called kernel





# References

- Fake News Detection Using Machine Learning Approaches. By Syed Ishfaq Manzoor, Dr Jimmy Singla , Prof. Nikita
- Evaluating Machine Learning Algorithms for Fake News Detection . By Prof. Shlok Glida.
- Detecting fake news in social media networks. By Monther Aldwairi , Ali Alwahedi.



Thank You !!

