

# CHAPTER 1

## INTRODUCTION

### 1.1 Background

Fake news is a type of hoax or deliberate spread of misinformation, be it via the traditional news media or via social media with the intent to mislead in order to gain financially or politically. (Hunt,E., 'What is Fake News ? How to spot it and what you can do to stop it', The Guardian, 17 December 2016; retrieved 25 September 2017)

In Indonesia, hoax news phenomena spread hate speech or strong emotion in their news content and social media. Indonesian Ministry of communications and information confirm there are 800 thousand sites in Indonesia that indicated spread hoax news and hate speech. ( Rudiantara, ' Ada 800 Ribu Situs Penyebar Hoax di Indonesia', CNN Indonesian News 29 December 2016). Based on Cambridge Dictionary, hate speech is a public speech that expresses hate or encourages violence towards a person or group based on something such as race, religion, sex or sexual orientation.

Hoax news has become powerful in Indonesia again since the Jakarta's governor election and the controversy about defamation that's already being made by candidate of Jakarta's governor lead many demonstrations such as 212 and destroy Indonesian politic environment . ( 'Seperti apa wajah politik Indonesia setelah Pilkada Jakarta berakhir?', BBC Indonesia 19 April 2017).

Based on that problem, that news has to be analyzed using sentiment analysis and classified into real or hoax news so the readers won't be easily being persuaded or misunderstood. Based on Oxford Dictionaries, sentiment analysis is the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc is positive, negative .

Text Mining is a process to gain textual information from data. Nazief & Adriani algorithm is one of text mining algorithm. This project will use this algorithm to gain textual information from on-line news.

Machine Learning is one area of artificial intelligence science that gives computer ability to learn from training data model without being explicitly programmed and suitable to implement sentiment analysis. This project also will use machine learning logic to classify the news into hoax or real news .

The multinomial Naive Bayes algorithm is one of machine learning algorithm. Naive Bayes will learn characteristic of hoax news based on training data model and classify test data from what already Naive Bayes learn before. With learning logic from machine learning, the news will be classified to the hoax or real news.

## **1.2 Scope**

This project is based on Java desktop and database mysql to store training data. The writer use arrays, tree and linked list to implement data from database toward Multinomial Naive Bayes algorithm and Nazief & Adriani Algorithm. Web scrapping and text input are facilities that provided to get test data & training data. Training real news data are retrieved from [tribunnews.com](http://tribunnews.com), [kompas.com](http://kompas.com), [detik.com](http://detik.com), [cnnindonesia.com](http://cnnindonesia.com). Hoax news are retrieved from [saracennews.com](http://saracennews.com), [seword.com](http://seword.com) and [is-hoax.blogspot.co.id](http://is-hoax.blogspot.co.id), [suaranasional.com](http://suaranasional.com). Nazief Adriani algorithm used to gain information from test data and training data. Machine learning logic will be used in this project so the result of test data will be training data. This project will focus on hate speech and strong emotion that contain in hoax news. Output of this project is conclusion, whether the news is categorized as hoax or real news.

Some questions that want to be proved in this project

1. How to get training data and test data using web scrapping and text input ?

2. How to gain information of data using Nazief & Adriani Algorithm ?
3. How to processing data using TF-IDF ?
4. How to implement Naive Bayes algorithm to classify the data into hoax or real news ?

### **1.3 Objective**

The first objective of this project is to classify the news or message with the sentiment analysis method in order to get relevant news and can't easily being persuaded by hoax news. The second objective of this project is to implement this project into desktop application that can be used easily when we want to read news from internet.

