AUTOMATED NEWS CATEGORIZER

MR. MARWAN YEEDENG

MR. NIHENG MAE

ADVISOR

MR. KHOLED LANGSAREE

FATONI UNIVERSITY

ABSTARCT

Nowadays, there are massive amount of data on the internet. The internet represents the contents delivery. People use internet to send data to one another. News agencies use deliver contents of news over the internet to people via web application. People read news over website a lot. News has been sent over the internet every day. With this massive amount of data will make a big problem that is to take a big amount of time to categorize. With this problem we found the solution is that to use machine learning to classify news automatically. Our project will use Naive Bayes machine learning to classify news. Once after classification. The data will be sent into SQL database. The output of the project is to visualize data on the website that we received from the database. The visualization will be in form of Bubbles visualization to visualize the amount of data.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1. OVERVIEW

Due to widespread availability of Internet, there are a huge resource that produce massive amount of daily news. Every day, people read a lot of news. Wither in social media like Twitter, Facebook or online news website. The news has been delivered to the people with different resources and even worse the news is represented differently or not talking about the same thing.

Because of this, it makes people being confused that what exactly happened lately? This makes a lot of people having no idea what are mostly people of the news throughout a month?

Unfortunately, the most of situations that happened in the country or in the world have been announced through news. Whenever people miss the news. They seem to be unprepared for the situations. This is why people should have known about what is the trend of the news in their daily life.

From the problem. We find out that it can be solved by using the technique of data mining to find out what is the trend of the news of each month? By using NLP (Natural Language Processing) to process the human language that is been written in news articles along with Naïve Bayes to categorize news into categories and summarize by using the technique of data visualization on a web application. The summarization will be about what is the trend of news of each month?

1.2. PROBLEM STATEMENTS

The main problem to the people who are reading online news are they cannot keep up with the news because there many online news sources or the internet and another problem is, they will miss up the news because there are variety of online news article on the internet like politics, economy, sport and so on. For those follow the only sport news they might up technology news or the others. Form those problem may cause the unexpected and unprepared situation problem. For instance, people have missed up the technology news. Once the technology changes the world. They will be unprepared for the situation.

From the problem above. The solution is having a news visualization to summaries the news form online news article. The output will be a web application that visualize trend of news of each month. The news will be shown in form of timeline visualization. The web application use news categorization with machine learning to categorize into its categories and visualize on a web application as a grouplike politics, sport, economy and so on.

1.3. OBJECTIVE OF RESEARCH

- To use the machine learning to categorize news and deployment by using technique of data visualization and data storytelling.
- To visualize the categorized news into news timeline visualization in form of data storytelling on a web application.

1.4. SCOPE

The scope of the project is to have a web application that summarize the trend of news of each month in form of data visualization the visualization will use date storytelling

technique to make beautiful data visualization and also the visualization can be interactive to users the form of visualization will can be a timeline of each month with describe the trend of news since the first data of the month to the last data of the month. The implementation of this project is useful for anyone. They will know the situation that are happening so that they will be prepared when the situation happened.

1.5. SIGNIFICANCE OF STUDY

1.5.1. DEVELOPER

- To gets technique skill of data analysis in term of using machine learning to apply with news categorizing.
- To deployment the result on web application by using python.

1.5.2. USER

- User can follow the trend of news are viral in social during that time.
- User gets quick access to relevant news and topics of interest.

1.6. SOFTWARE AND HARDWARE ARE REQUIREMENT

1.6.1.1. SOFTWARE REQUIREMENT

1.6.1.1.1. Resource

- News from Bangkok post agency website.

1.6.1.1.2. Data Mining Process

- Python

- Jupyter Notebook
- Data Mining Libraries
    - Requests
    - Beautiful Soup 4
    - Pandas
    - NumPy
    - re (Regular expression)

1.6.1.1.3. Data Visualization Process

- Website Language
    - HTML
    - JavaScript
    - CSS
    - Python
- Data Visualization Libraries
    - Django
    - Matplotlib
    - Seaborn

1.6.1.1.4. Data Base

- MySQL
- Data Integration Format
    - JSON

1.6.1.2. HARDWARE REQUIREME

1.6.1.2.1. Personal Computer

- Asus VivoBook 15 x512da
- HP Pavilion Power 15-cb035TX

CHAPTER II

LITERATURE REVIEW

Automated news categorization is news categorizer that use machine learning to categorize news and deploy as data virtualization on a website.

- Definition.
- Machine Learning.
- Information of software/hardware used in development of a system.
- Related works.

2.1. DEFINITION

2.1.1. NEWS CATEGORIZATION

The process of grouping news into its categories, for example sport news should be in a group of news about sport or politics news should be in a group of politics news.

2.1.2. AUTOMATED NEWS CATEGOIRZER

News categorizer that uses machine learning algorithm like classification to categorize news into its categories automatically. The machine learning model will be naive bayes to be trained on training data set that is news articles from online news.

2.1.3. DATA VISUALIZATION

The process that is used to visualize data. This process will be applied to illustrate the data that is news categories that has been categorized by machine learning. The project will use technique of data visualization to summarize the data of news and deploy on a web application.

2.2. MACHINE LEARNING

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

2.1.4. CLASSIFICATION

Classification is supervised algorithm that learn from the given data to make new classification into its group. The project will use classification algorithm to classify news into its categories.

2.1.5. NAÏVE BAYES CLASSIFIER

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e., every pair of features being classified is independent of each other.

2.3. TOOLS

2.1.6. CODE EDITOR

2.1.6.1. Jupyter Notebook

Jupyter Notebook is one of the most popular tools in the field of data science, which involves a lot of data management work. And still have to report research. Jupyter Notebook has been designed to meet the purpose of use, whether it is Access the library and write the code and see the results. Jupyter Notebook is designed to be more functional and readable than a normal program.

2.1.7. DATA MINING LIBRARIES

The libraries that help That control in the field of data management. Used to retrieve data from website and modify data.

2.1.7.1. Requests

Requests is a Python HTTP library, released under the Apache License 2.0. The goal of the project is to make HTTP requests simpler and more human-friendly. The current version is 2.25.0

2.1.7.2. Beautiful Soup 4

Beautiful Soup is a library that makes it easy to scrape information from web pages. It sits atop an HTML or XML parser, providing Pythonic idioms for iterating, searching, and modifying the parse tree.

2.1.7.3. Numpy

NumPy is a Python library that provides a simple yet powerful data structure: the n-dimensional array. This is the foundation on which almost all the power of Python's data science toolkit is built.

2.1.8. MACHINE LEARING

2.1.8.1. SCIKIT-LEARN

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.

2.1.9. DATA VISUALIZATION

The tools that are used to visualize data.

2.1.9.1. DJANGO

A Python-based free and open-source web framework that follows the model-template-views (MTV) architectural pattern. The purpose of this frame work is to build website with python.

2.1.9.2. MATPLOTLIB

A plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots

into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.

2.1.9.3. SEABORN

A Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

2.4. RELATE WORKS

2.1.10. AUTOMATIC SEMANTIC CATEGORIZATION OF NEWS HEADLINES USING ENSEMBLE MACHINE LEARNING: A COMPARATIVE STUDY.

Due to widespread availability of Internet, there are a huge of sources that produce massive amounts of daily news. Moreover, the need for information by users has been increasing unprecedently, so it is critical that the news is automatically classified to permit users to access the required news instantly and effectively. One of the major problems with online news sets is the categorization of the vast number news and articles. In order to solve this problem, the machine learning model along with the Natural Language Processing (NLP) is widely used for automatic news classification to categorize topics of untracked news and individual opinion based on the user's prior interests. However, the existing studies mostly rely on NLP but uses huge documents to train the prediction model, thus it is hard to classify a short text without using semantics. Few studies focus on exploring classifying the news headlines using the semantics. Therefore, this paper attempts to use semantics and ensemble learning to improve the short text classification. The proposed

methodology starts with preprocessing stage then applying feature engineering using word2vec with TF-IDF vectorizer. Afterwards, the classification model was developed with different classifier KNN, SVM, Naïve Bayes and Gradient boosting. The experimental results verify that Multinomial Naïve Bayes shows the best performance with an accuracy of 90.12% and recall 90%. (Bogery et al., 2019)

2.1.11. A COMPARATIVE ANALYSIS OF NEWS CATEGORIZATION USING MACHINE LEARNING APPROACHES.

The rapid growth of print and digital media increased the reach of one and all in terms of information, resulting in more amount of Text data to be mined. This data is nothing but a heap of unclassified information which when kept together means nothing. This means that there is a need to tag all these data i.e., News Classification. News classification is the task of automatically classify the news documents into their predefined classes based on their content with the confidence learned from the training news dataset. This research evaluates some most widely used machine learning techniques, mainly Naive Bayes, Random Forest, Decision Tree, SVM and Neural Networks, for automatic news classification problem. To experiment the system, a dataset from BBC that have two columns, one has the news headlines and the other contains the type it belongs to. There are 2225 rows in the data set is used. The average results show that the Naive Bayes is performing better than the other four algorithms with the classification accuracy of 96.8 % .Then follows the Random Forest with accuracy 94.1 % , Support Vector Machine (SVM) with accuracy 96.4 %, Neural Networks with accuracy 96.4 % and the Decision Tree with accuracy 83.2 %. (Deb et al., 2020)

2.1.12. TEXT MINING APPROACH TO CLASSIFY TECHNICAL RESEARCH DOCUMENT USING NAÏVE BAYES

World Wide Web is the store house of abundant information available in various electronic forms. Since past few years, the increase in the performance of computers in handling large quantity of text data has led researchers to focus on reliable and optimal retrieval of visible and implied information that exist in the huge resources. In text mining, one of the challenging and growing importance's is given to the task of document classification or text characterization. In this process, reliable text extraction, robust methodologies and efficient algorithms such as Naive Bayes and other made the task of document classification to perform consistently well. Classifying text documents using Bayesian classifiers are among the most successful known algorithms for machine learning. This paper describes implementations of Naïve Bayesian (NB) approach for the automatic classification of Documents restricted to Technical Research documents based on their text contents and its results analysis. We also discuss a comparative analysis of Weighted Bayesian classifier approach with the Naive Bayes classifier. (M et al., 2015)

CHAPTER III

METHODOLOGY

This chapter focuses on methodology of project automated news categorizer by following steps of CRISP-DM process model.

3.1. INTRODUCTION

3.1.1. DEFINITION OF CRISP-DM

The CRoss Industry Standard Process for Data Mining (CRISP-DM) is a process model with six phases that naturally describes the data science life cycle. It's like a set of guardrails to help you plan, organize, and implement your data science (or machine learning) project. (Saltz & Hotz, 2020)

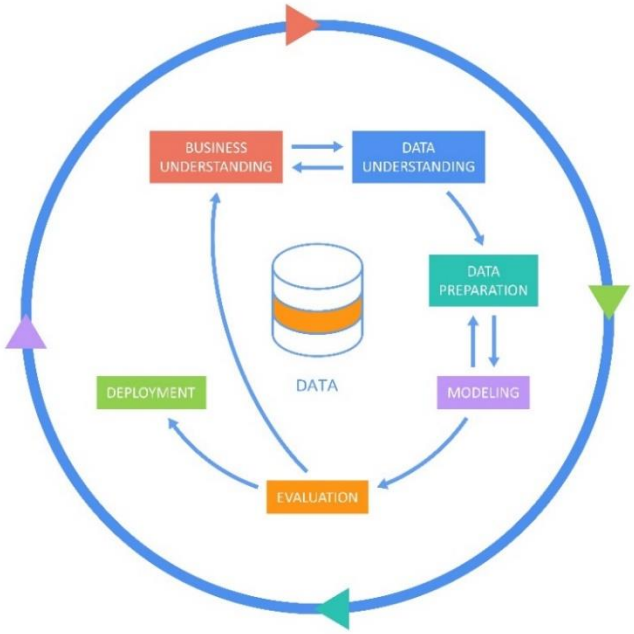


Figure 3.1: CRoss Industry Standard Process for Data Mining (CRISP-DM)

3.2. CRISP-DM

3.1.2. BUSINESS UNDERSTADING

Business Understanding phase, focuses on understanding the objectives and requirements of the project. Aside from the third task, the three other tasks in this phase are foundational project management activities that are universal to most projects.

3.1.2.1. Determine Business Objectives

To have a web application that summarize the trend of news of each month in form of data visualization the visualization will use date storytelling technique to make beautiful data visualization and also the visualization can be interactive to users the form of visualization will can be a timeline of each month with describe the trend of news since the first data of the month to the last data of the month. The implementation of this project is useful for anyone. They will know the situation that are happening so that they will be prepared when the situation happened.

3.1.2.2. Assess Situation

The Bangkok post agency their website has enough the necessary resource for to do the categorizing news and visualization, the website their using the English language that easy and support to preparing the data, training model and also testing model.

3.1.2.3. Determine Data Mining Goals

News categorizer to categorize news into its category.

3.1.2.4. Produce Project Plan

Business Under-standing	Categorizing news automatically by using machine learning Naïve Bayes.
Data Understanding	**Resource** Bangkok post agency website. **File Extension** JSON.
Data Preparation	NLP (Natural Language Processing).
Modeling	Naïve Bayes.
Evaluation	Model Evaluation.
Deployment	Web Application.

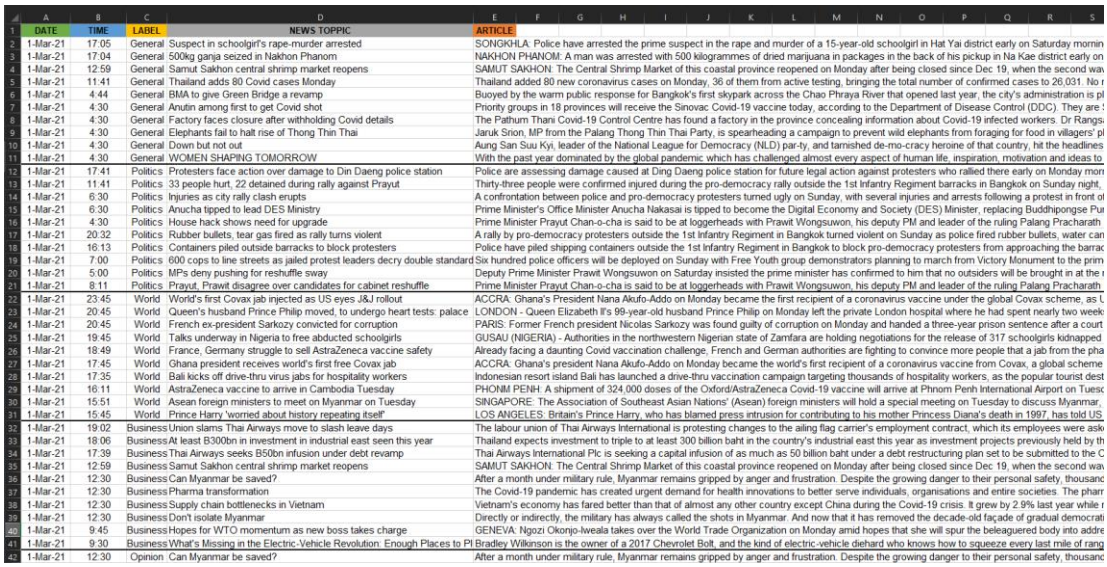
Table 3.2.1.4: Produce Project Plan

3.1.3. DATA UNDERSTANDING

Data Understanding phase, adding to the foundation of Business Understanding, it drives the focus to identify, collect, and analyze the data sets that can help accomplish the project goals. This phase also has four tasks.

3.1.3.1. Collect Initial Data

Collecting the necessary data for to do categorization and visualization, the necessary data for this project are Date of news publish, Time of news publish, News Category, News topic, News article.



	A	B	C	D	E
1	DATE	TIME	LABEL	NEWS TOPPIC	ARTICLE
2	1-Mar-21	17:05	General	Suspect in schoolgirl's rape-murder arrested	SONGKHLA: Police have arrested the prime suspect in the rape and murder of a 15-year-old schoolgirl in Hat Yai district early on Saturday mornin
3	1-Mar-21	17:04	General	500kg ganja seized in Nakhon Phanom	NAKHON PHANOM: A man was arrested with 500 kilogrammes of dried marijuana in packages in the back of his pickup in Na Kae district early on
4	1-Mar-21	12:59	General	Samut Sakhon central shrimp market reopens	SAMUT SAKHON: The Central Shrimp Market of this coastal province reopened on Monday after being closed since Dec 19, when the second wav
5	1-Mar-21	11:41	General	Thailand adds 80 Covid cases Monday	Thailand added 80 new coronavirus cases on Monday, 36 of them from active testing, bringing the total number of confirmed cases to 26,031. No r
6	1-Mar-21	4:44	General	BMA to give Green Bridge a revamp	Buoyed by the warm public response for Bangkok's first skypark across the Chao Phraya River that opened last year, the city's administration is pl
7	1-Mar-21	4:30	General	Anutin among first to get Covid shot	Priority groups in 18 provinces will receive the Sinovac Covid-19 vaccine today, according to the Department of Disease Control (DDC). They are :
8	1-Mar-21	4:30	General	Factory faces closure after withholding Covid details	The Pathum Thani Covid-19 Control Centre has found a factory in the province concealing information about Covid-19 infected workers. Dr Rangsi
9	1-Mar-21	4:30	General	Elephants fail to halt rise of Thong Thin Thai	Jaruk Srion, MP from the Palang Thong Thin Thai Party, is spearheading a campaign to prevent wild elephants from foraging for food in villagers' pl
10	1-Mar-21	4:30	General	Down but not out	Aung San Suu Kyi, leader of the National League for Democracy (NLD) par-ty, and tarnished de-mo-cracy heroine of that country, hit the headlines
11	1-Mar-21	4:30	General	WOMEN SHAPING TOMORROW	With the past year dominated by the global pandemic which has challenged almost every aspect of human life, inspiration, motivation and ideas to
12	1-Mar-21	17:41	Politics	Protesters face action over damage to Din Daeng police station	Police are assessing damage caused at Ding Daeng police station for future legal action against protesters who rallied there early on Monday morn
13	1-Mar-21	11:41	Politics	33 people hurt, 22 detained during rally against Prayut	Thirty-three people were confirmed injured during the pro-democracy rally outside the 1st Infantry Regiment barracks in Bangkok on Sunday night,
14	1-Mar-21	6:30	Politics	Injuries as city rally clash erupts	A confrontation between police and pro-democracy protesters turned ugly on Sunday, with several injuries and arrests following a protest in front of
15	1-Mar-21	6:30	Politics	Anucha tipped to lead DES Ministry	Prime Minister's Office Minister Anucha Nakasai is tipped to become the Digital Economy and Society (DES) Minister, replacing Buddhipongse Pur
16	1-Mar-21	4:30	Politics	House hack shows need for upgrade	Prime Minister Prayut Chan-o-cha is said to be at loggerheads with Prawit Wongsuwon, his deputy PM and leader of the ruling Palang Pracharath
17	1-Mar-21	20:32	Politics	Rubber bullets, tear gas fired as rally turns violent	A rally by pro-democracy protesters outside the 1st Infantry Regiment in Bangkok turned violent on Sunday as police fired rubber bullets, water can
18	1-Mar-21	16:13	Politics	Containers piled outside barracks to block protesters	Police have piled shipping containers outside the 1st Infantry Regiment in Bangkok to block pro-democracy protesters from approaching the barrac
19	1-Mar-21	7:00	Politics	600 cops to line streets as jailed protest leaders decry double standard	Six hundred police officers will be deployed on Sunday with Free Youth group demonstrators planning to march from Victory Monument to the prim
20	1-Mar-21	5:00	Politics	MPs deny pushing for reshuffle sway	Deputy Prime Minister Prawit Wongsuwon on Saturday insisted the prime minister has confirmed to him that no outsiders will be brought in at the r
21	1-Mar-21	8:11	Politics	Prayut, Prawit disagree over candidates for cabinet reshuffle	Prime Minister Prayut Chan-o-cha is said to be at loggerheads with Prawit Wongsuwon, his deputy PM and leader of the ruling Palang Pracharath.
22	1-Mar-21	23:45	World	World's first Covax jab injected as US eyes J&J rollout	ACCRA: Ghana's President Nana Akufo-Addo on Monday became the first recipient of a coronavirus vaccine under the global Covax scheme, as U
23	1-Mar-21	20:45	World	Queen's husband Prince Philip moved, to undergo heart tests: palace	LONDON - Queen Elizabeth II's 99-year-old husband Prince Philip on Monday left the private London hospital where he had spent nearly two week
24	1-Mar-21	20:45	World	French ex-president Sarkozy convicted for corruption	PARIS: Former French president Nicolas Sarkozy was found guilty of corruption on Monday and handed a three-year prison sentence after a court
25	1-Mar-21	19:45	World	Talks underway in Nigeria to free abducted schoolgirls	GUSAU (NIGERIA) - Authorities in the northwestern Nigerian state of Zamfara are holding negotiations for the release of 317 schoolgirls kidnapped
26	1-Mar-21	18:49	World	France, Germany struggle to sell AstraZeneca vaccine safety	Already facing a daunting Covid vaccination challenge, French and German authorities are fighting to convince more people that a jab from the pha
27	1-Mar-21	17:45	World	Ghana president receives world's first free Covax jab	ACCRA: Ghana's president Nana Akufo-Addo on Monday became the world's first recipient of a coronavirus vaccine from Covax, a global scheme
28	1-Mar-21	17:35	World	Bali kicks off drive-thru virus jabs for hospitality workers	Indonesian resort island Bali has launched a drive-thru vaccination campaign targeting thousands of hospitality workers, as the popular tourist dest
29	1-Mar-21	16:11	World	AstraZeneca vaccine to arrive in Cambodia Tuesday	PHONM PENH: A shipment of 324,000 doses of the Oxford/AstraZeneca Covid-19 vaccine will arrive at Phnom Penh International Airport on Tuesc
30	1-Mar-21	15:51	World	Asean foreign ministers to meet on Myanmar on Tuesday	SINGAPORE: The Association of Southeast Asian Nations' (Asean) foreign ministers will hold a special meeting on Tuesday to discuss Myanmar,
31	1-Mar-21	15:45	World	Prince Harry 'worried about history repeating itself'	LOS ANGELES: Britain's Prince Harry, who has blamed press intrusion for contributing to his mother Princess Diana's death in 1997, has told US
32	1-Mar-21	19:02	Business	Union slams Thai Airways move to slash leave days	The labour union of Thai Airways International is protesting changes to the ailing flag carrier's employment contract, which its employees were aske
33	1-Mar-21	18:06	Business	At least B300bn in investment in industrial east seen this year	Thailand expects investment to triple to at least 300 billion baht in the country's industrial east this year as investment projects previously held by th
34	1-Mar-21	17:39	Business	Thai Airways seeks B50bn infusion under debt revamp	Thai Airways International Plc is seeking a capital infusion of as much as 50 billion baht under a debt restructuring plan set to be submitted to the C
35	1-Mar-21	12:59	Business	Samut Sakhon central shrimp market reopens	SAMUT SAKHON: The Central Shrimp Market of this coastal province reopened on Monday after being closed since Dec 19, when the second wav
36	1-Mar-21	12:30	Business	Can Myanmar be saved?	After a month under military rule, Myanmar remains gripped by anger and frustration. Despite the growing danger to their personal safety, thousanc
37	1-Mar-21	12:30	Business	Pharma transformation	The Covid-19 pandemic has created urgent demand for health innovations to better serve individuals, organisations and entire societies. The pharr
38	1-Mar-21	12:30	Business	Supply chain bottlenecks in Vietnam	Vietnam's economy has fared better than that of almost any other country except China during the Covid-19 crisis. It grew by 2.9% last year while r
39	1-Mar-21	12:30	Business	Don't isolate Myanmar	Directly or indirectly, the military has always called the shots in Myanmar. And now that it has removed the decade-old façade of gradual democrati
40	1-Mar-21	9:45	Business	Hopes for WTO momentum as new boss takes charge	GENEVA: Ngozi Okonjo-Iweala takes over the World Trade Organization on Monday amid hopes that she will spur the beleaguered body into addre
41	1-Mar-21	9:30	Business	What's Missing in the Electric-Vehicle Revolution: Enough Places to Pl	Bradley Wilkinson is the owner of a 2017 Chevrolet Bolt, and the kind of electric-vehicle diehard who knows how to squeeze every last mile of rang
42	1-Mar-21	12:30	Opinion	Can Myanmar be saved?	After a month under military rule, Myanmar remains gripped by anger and frustration. Despite the growing danger to their personal safety, thousanc

Figure 3.2.2.1: Data set of Automated news categorizer

3.1.3.2. Describe Data

File Extension	JSON
Columns	6 Columns
Rows	3,000 Rows

Table 3.2.2.2: Describe Data

3.1.3.3. Explore Data

Variable	Definition	Data Type	Objective
Date	The day of news publish.	Datetime	To determine the timeline of news.
Time	The time of news publish.	Datetime	To determine the timeline of news.
Label	The type of news category.	Str	To train super-vised machine learning.
News topic	The topic of news.	Str	To determine the topic news.
Article	News article.	Str	To train super-vised machine learning.

Table 3.2.2.3: Explore Data

3.1.3.4. Verify Data Quality

Checking the data set that which column and row are going to clean, the result of data set of automated news categorizer the article column that have punctua-tion marks.

3.1.4. DATA PREPARATION

Data Preparation phase, which is often referred to as “data munging”, prepares the final data set for modeling. It has five tasks.

3.1.4.1. Select Data

Variable	Definition	Data Type	Objective
Label	The type of news category.	Str	Determine the category of news to train supervised machine learning.
Article	News article.	Str	To train supervised machine learning.

Table 3.2.3.1: Select Data

3.1.4.2. Clean Data

Cleaning the article column by using NLP (Natural Language Processing) remove the punctuation marks.

3.1.5. MODELING

Modeling phase, Select the modeling techniques are suitable with project. It has four tasks.

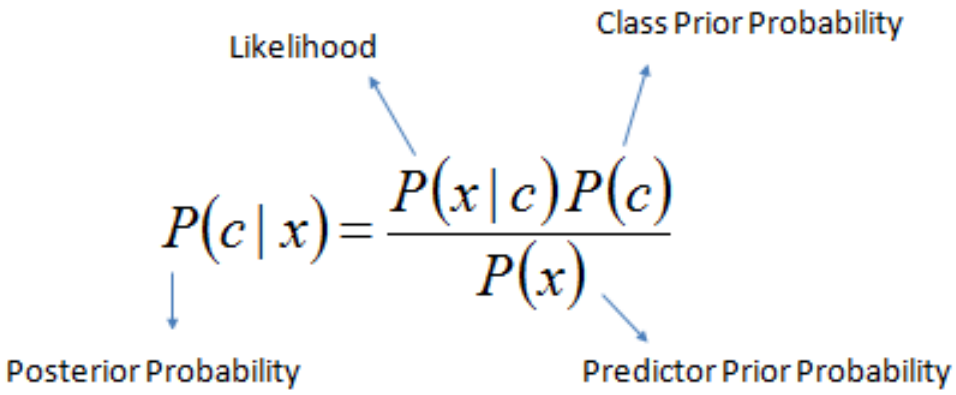
3.1.5.1. Select Modeling Techniques

Using Naïve Bayes classifier model from scikit-learn to classifies a news.

3.1.5.2. Generate Test Design

Using the article column 3,000 record to training the model and testing the model by speared 70% training and 30% testing.

3.1.5.3. Build Model



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Figure 3.2.4.3: Naïve Bayes algorithm (Sayad, 2010)

3.1.5.4. Assess Model

Naïve Bayes classifier model to assess Output, Accuracy, Time complexity, Error.

3.1.6. EVALUATION

Evaluation phase, looks more broadly at which model best meets the business and what to do next. This phase has three tasks.

3.1.6.1. Evaluate Results

Evaluate the results of training model and testing model, do the model meet the business understanding success or not by using model evaluation.

3.1.6.2. Review Process

Review whole process to see if there are any parts to be missed.

3.1.6.3. Determine Next Steps

To retrieve the result, deployment on the web application as data visualization in form of news timeline.

3.1.7. DEPLOYMENT

Deployment phase, the model is not particularly useful unless apply it with the real system that following the goal of Business Understanding. The complexity of this phase varies widely. This final phase has four tasks.

3.1.7.1. Plan Deployment

Build a web application for deployment. The visualization will be in form of news timeline.

REFERENCES

Bogery, R., Al Babtain, N., Aslam, N., Alkabour, N., Al Hashim, Y., & Ullah Khan, I. (2019). Automatic Semantic Categorization of News Headlines using Ensemble Machine Learning: A Comparative Study. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 10, Issue 11). www.ijacsa.thesai.org

Deb, N., Jha, V., Panjiyar, A. K., & Gupta, R. K. (2020). A comparative analysis of news categorization using machine learning approaches. *International Journal of Scientific and Technology Research*, *9*(1), 2469–2472. www.ijstr.org

M, M. K., H, S. D., Desai, P. G., & Chiplunkar, N. (2015). Text Mining Approach to Classify Technical Research Documents using Naïve Bayes. *International Journal of Advanced Research in Computer and Communication Engineering*, *4*(7), 386–391. https://doi.org/10.17148/IJARCCE.2015.4789

Saltz, J., & Hotz, N. J. (2020). *CRISP-DM Data Science Project Management*. Https://Www.Datascience-Pm.Com/Crisp-Dm-2/. https://www.datascience-pm.com/crisp-dm-2/

Sayad, D. S. (2010). *Naive Bayesian*. 1–4. https://www.saedsayad.com/naive_bayesian.htm