Trends of News Visualization Project

Mr.Marwan Yeedeng

Mr.Niheng   Mae

Adviser

Mr.Kholid Langsaree

Overview

The project of our team is about data visualization of Thailand News Trends Over 4 Years. We will use the study of Data mining to analyze the News of Top 10 Thailand News Agency Websites to answer the question of ours that is ***"What is the Trends of Thailand News Over 4 Years?"*** and the output of the research will be exported by using the technique of data storytelling as Data Timeline Visualization.

We will use website technology to deploy our product because the website will make our data having interaction with users and look more beautiful than just a static graph.

Problem Statements

The project is to the answer of the question to answer the question ***"What is the Trends of Thailand News Over 4 Years?"*** but to answer the question we need to have the data. To collect the needed data and output those as the answer of the question we need to use the field of study of data mining and its process like crisp-dm.

Objective of Research

- We would like to use the study of data mining to analyze the trend of news of Thai News Agencies.
- We would like to use the study of data storytelling to illustrate the trend of news of Thai News Agencies.

Scope

The scope of this proposed project is to use techniques of Data Mining technology to summarize the Trends of News and deploy as data visualization in from of Timeline.

Significance of Study

- The question of the research will be answered.
- Improve the field of study of data mining knowledge.
- Expand the body of knowledge.

Software and Hardware Requirement

Resource

- News from 10 Thailand News Agency Websites.

Data Mining Process

- Python
- Jupyter notebook

- Data Mining Libraries
  - Pandas
  - Numpy
  - Requests

Data Visualization Process

- Web Application Languages
  - Html
  - JavaScript
  - Css
- Data Visualization Libraries
  - Matplotlib
  - Seaborn

Database

- Spark
- Data Integration Format
  - Json

Methodology

Planning: During planning the project. We will use Crisp-Dm to plan the discipline of the project. Crisp-DM or Cross Industry Standard Process for Data Mining is for planning the one of data-mining process that used to identify the step of doing data mining project.

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
• Business Objectives	• Collect Initial Data	• Select Data	• Select Modeling Techniques	• Evaluate Result	• Plan Deployment
• Assess Situation	• Describe Data	• Clean Data	• Generate Test Design	• Review Process	• Plan Monitoring and Maintenance
• Data Mining Goals	• Explore Data	• Construct Data	• Build Model	• Determine Next Steps	• Product Final Report
• Produce Project Plan	• Verify Data Quality	• Integrate Data	• Assess Model		• Review Project
		• Format Data			

Processing: We will use Anaconda jupyter notebook as text-editor with Python Programming Language. Start with data acquisition by scraping from Thailand News Agencies and hen use python data science library like pandas to clean them up as a useful data that we can take and use and also summarize in term of statistical mathematics by using python and numpy.

Deployment: In the process of Deployment. We will deploy as statistical mathematics by using pandas and numpy. Statistical graph by using matplotlib and seaborn. Final step is using building a website to deploy our project to make it more interaction to users and using the technique of data storytelling tools to illustrate it.