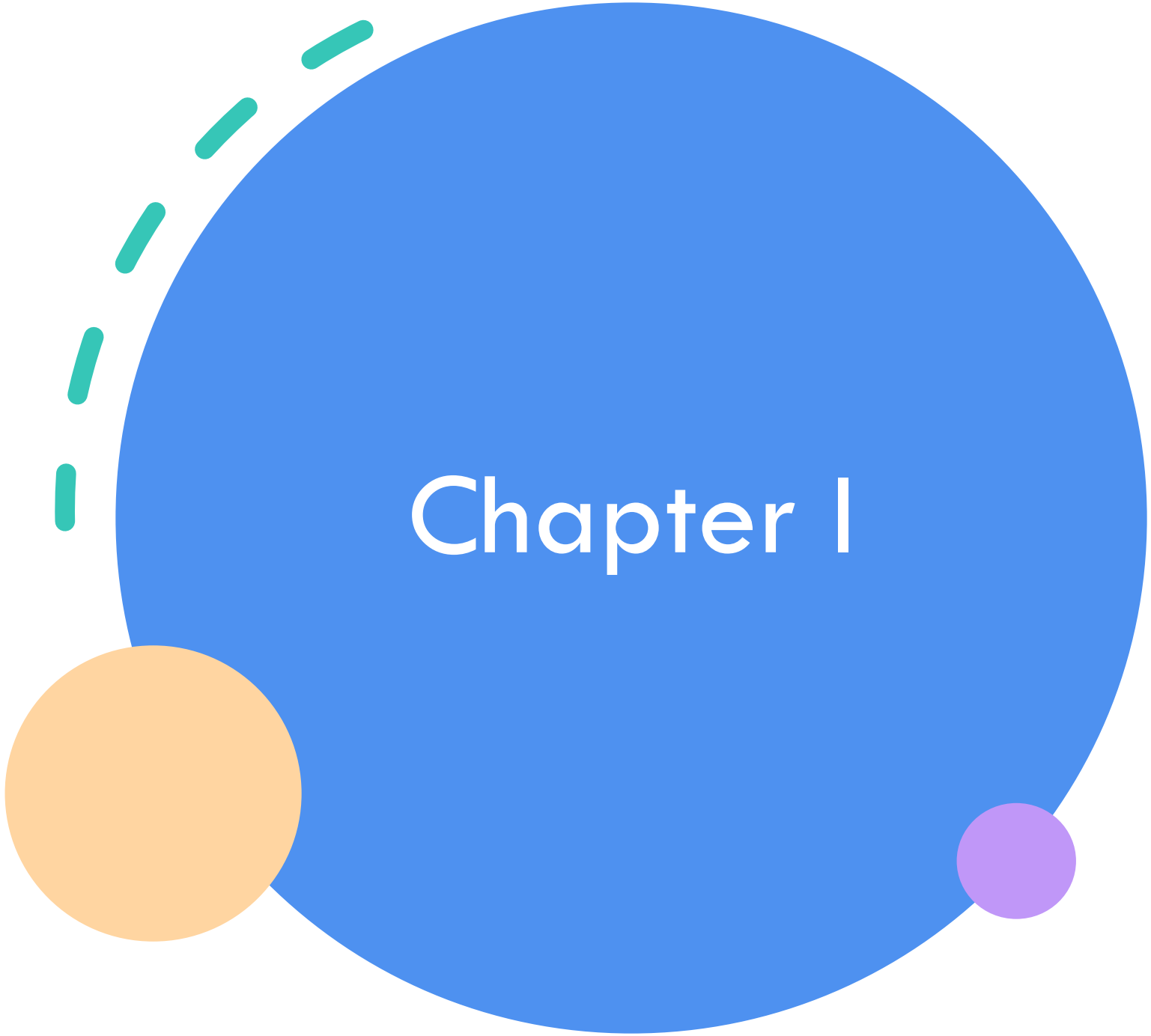
Automated News Categorizer

Proposal Three Chapter Presentation

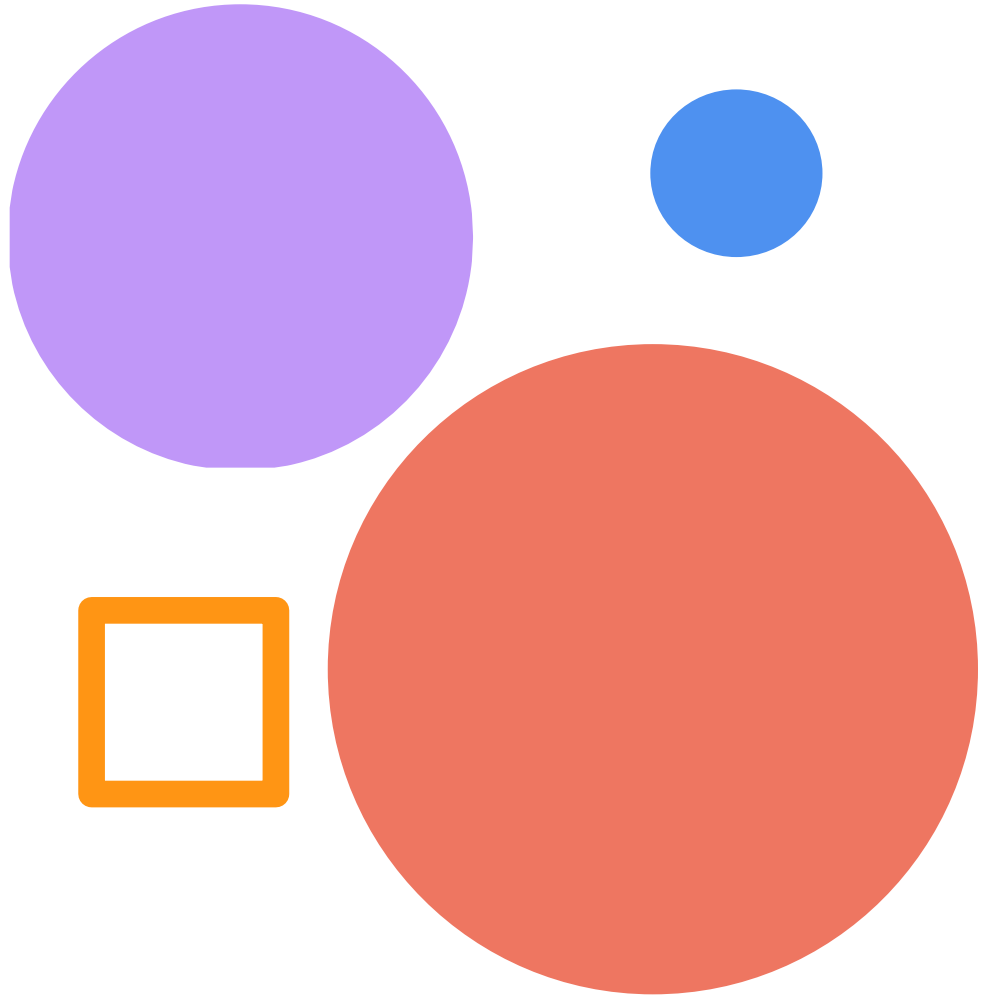
Content


Chapter 1

Content

- Chapter I
  - Overview.
  - Problem Statement.
  - Objective of Research.
  - Scope.
  - Significance of Study.
  - Software and Hardware Are Requirement.

Overview

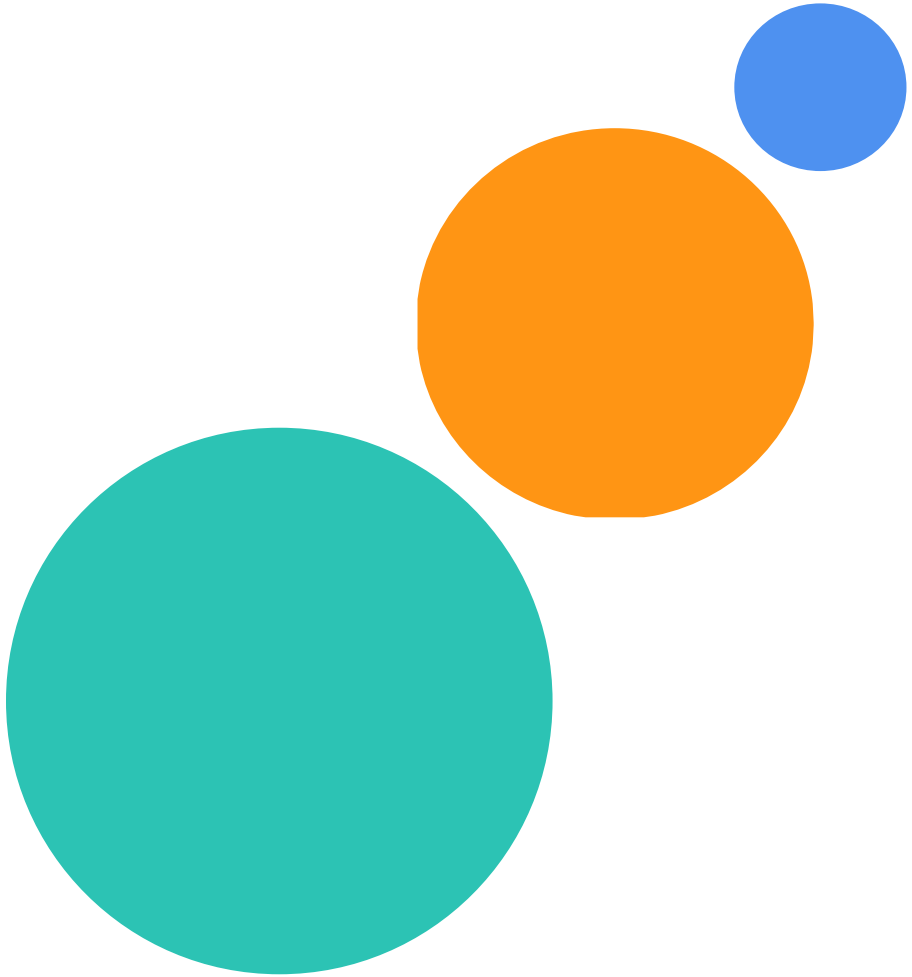


Problem Statement

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.

Scope



Significance of Study

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.

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.

Software Requirement

Resource

- News from Bangkok post agency website.

Data Mining Process

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

Software Requirement

Data Visualization Process

- Web Application Language
  - HTML
  - JavaScript
  - Css
  - Python
- Data Visualization Libraries
  - Django
  - Matplotlib
  - Seaborn

Data Base

- MySQL
- Data Integration Format
  - JSON

Hardware Requirement

Personal Computer

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

Chapter II

Content

- Chapter II
  - Definition.
  - Machine learning.
  - Infor of software/hardware in dev of system.
  - Relate works

Chapter II

Code Editor

- Jupyter Notebook.

Machine Learning

- Classification.
- Naïve Bayes Classifier.

Tool

Chapter II

Definition

- News Categorization.
- Automated News Categorizer.
- Data Visualization.

Data Mining Libraries

- Requests.
- Beautiful Soup 4.
- Numpy.

Chapter II

Machine Learning

- Scikit-Learn

Relate works

Chapter II

- Automated Semantic Categorization of News Headlines Using Ensemble Machine Learning: A Comparative Study. (Bogery et al., 2019)

- A COMPARATIVE ANALYSIS OF NEWS CATEGORIZATION USING MACHINE LEARNING APPROACHES. (Deb et al., 2020)

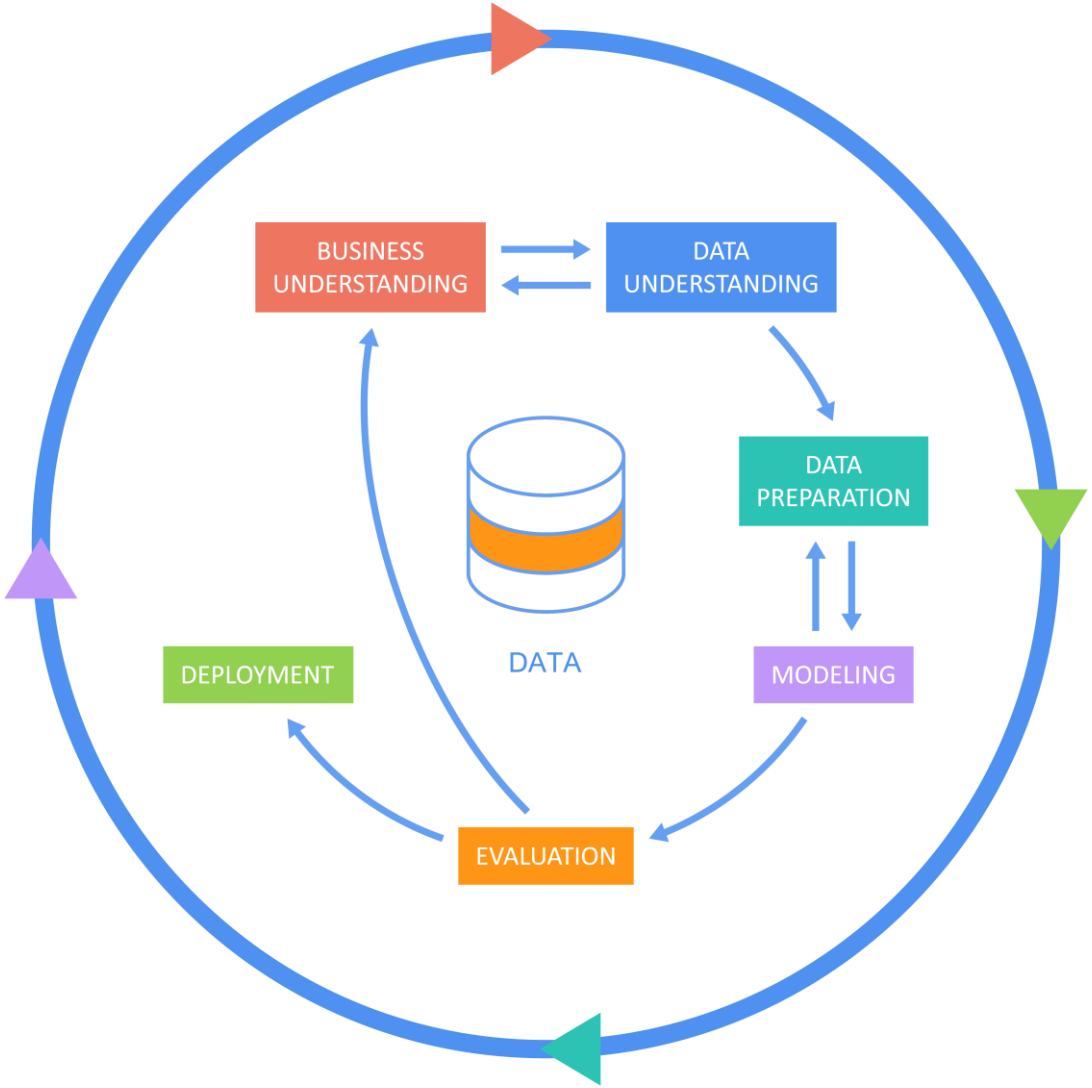
Chapter III

Content

- Chapter III
  - CRISP-DM

Methodology

CRISP-DM



Methodology

Business Understanding

3.2.1.1. Determine Business Objectives.

3.2.1.2. Assess Situation.

3.2.1.3. Determine Data Mining Goals.

3.2.1.4. Produce Project Plan.

Methodology

Data Understanding

3.2.2.1. Collect Initial Data.

3.2.2.2. Describe Data.

3.2.2.3. Explore Data.

3.2.2.4. Verify Data Quality.

Methodology

Data Preparation

3.2.3.1. Select Data.

3.2.3.2. Clean Data.

Methodology

Modeling

3.2.4.1. Select Modeling Techniques.

3.2.4.2. Generate Test Design.

3.2.4.3. Build Model.

3.2.4.4. Assess Model.

Methodology

Evaluation

3.2.5.1. Evaluate Results.

3.2.5.2. Review Process.

3.2.5.3. Determine Next Steps.

Methodology

Deployment

3.2.6.1. Plan Deployment.

Thank You :)