

## CHAPTER 3

### RESEARCH METHODOLOGY

#### 3.1. Research Framework

The flows of this research basically, follow the guidelines of the existing Kansei engineering application and text mining integration by (Hsia, Chen and Lin 2017), that in this research applies a different approach on opinion mining focused on LDA topic modelling, the detailed modified flows of this research can be seen in Figure 3.1.

#### 3.2. Literature Study

Literature reference is used as the basis of thoughts in this research. Empirical references that use the same methods and techniques with this research are studied to gain a new knowledge on the position of this research towards the application of these technique. And other than that, theoretical literature study is conducted to get the idea of the basis of the theories that are applied here.

### **3.3. Data Collection**

This research retrieves the dataset of Samsung S9 reviews from amazon.com via web-scraping method in python, secondly, the other smartphone reviews dataset is taken and downloaded from a dataset website named Kaggle.com. The other smartphone brand dataset which contains the review about various smartphone brand such as Apple, Nokia, LG, Motorola, Lenovo, and even the older Samsung phone like galaxy S6 until galaxy S8, which later to be used as the comparison for the first dataset about Samsung S9 reviews.

### **3.4. Feature-Based Opinion Mining**

After the data collection process, the next step is to conduct a data analysis which is done in NLP packages in python, namely feature-based opinion mining that implements LDA topic modelling, which consist of several sub-processes. It is text pre-processing, that deals with changing the semi-structured customer reviews data text into a more readable for python to be processed later, phrase modelling that combines n-grams of words inside the dataset, then for the last step is implementation of LDA Topic modelling that use gensim library in Python. After all the mentioned processes are finished, the next step is to interpret the result of LDA Topic modelling itself.

### **3.5. Features Identification**

After getting the interpretation result of LDA, we can see some of the topics formed up the most probable highlighted topics from the whole reviews. The both datasets contain a different highlighted feature, so we hope these insights will give us a good result.

### **3.6. Kansei Words Identification**

This process also derived from the result of LDA, that the words only consist of adjectives from the reviews that indicate the impressions of customers towards the product, especially towards the features of the products. On both datasets will also contain a different Kansei words. This process uses gensim word2vec library in python that calculates the cosine similarities of reviewed product's features and the Kansei words. By analysing the cosine similarities on each gathered feature from the products, it can be seen the relation between Kansei words with the features.

### **3.7. Kansei Words Matching and Comparison**

After the features being highlighted by LDA and its Kansei words, a comparison analysis is done to filter the Kansei words and its related features from reviews of Samsung S9 other smartphones. The idea of these comparison is to get a better insight of the Kansei words to be used in product improvement guideline.

### **3.8. Product Improvement Guideline**

This process builds up the summary on the findings of the features and Kansei words relations, that it can be used as a guideline for a future product improvement.

### **3.9. Conclusion and Recommendations**

This section will briefly answer the question stated in problem formulation formulated in chapter 1. Besides, this section also gives some recommendations that might be used to further research.

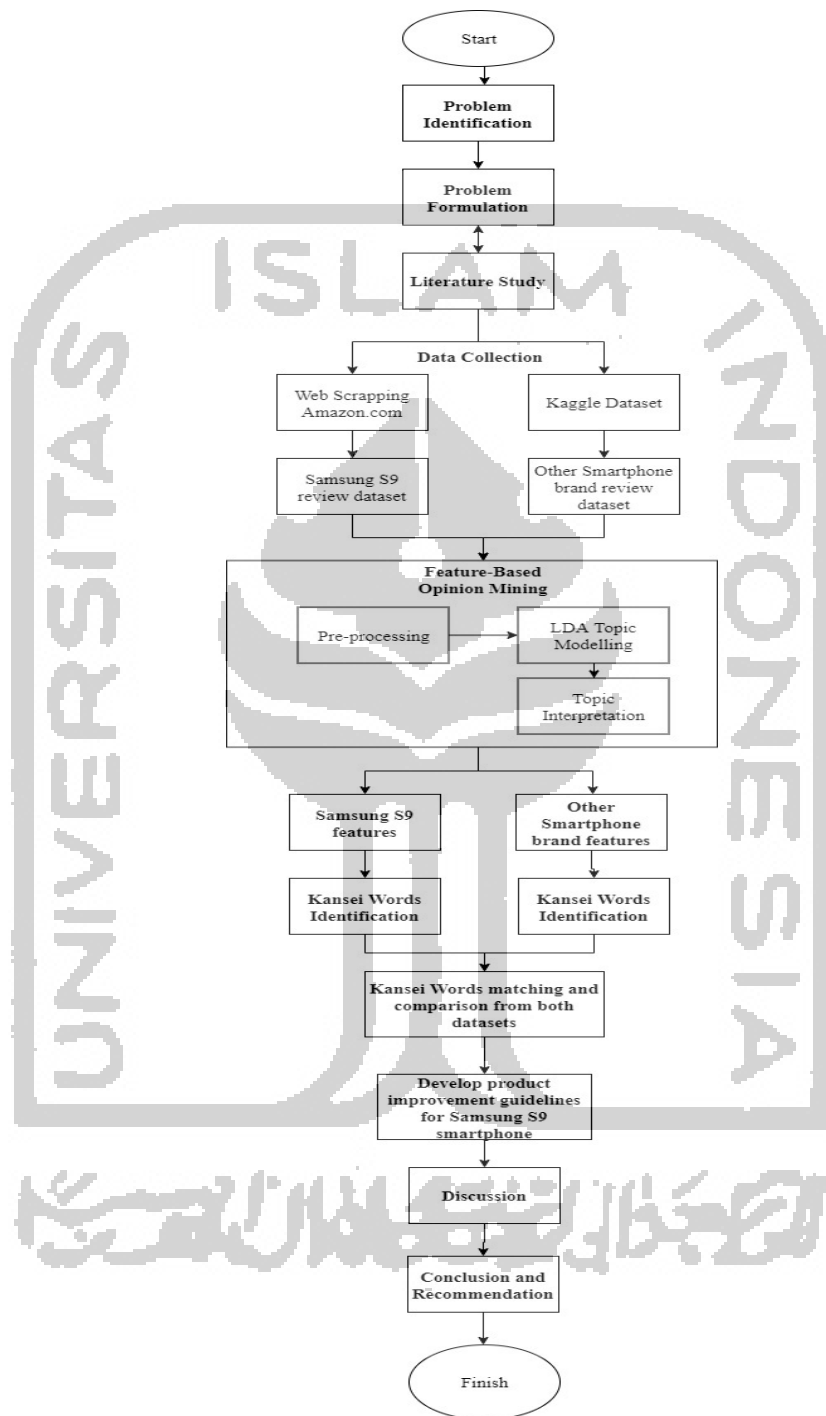


Figure 3. 1. Research Flowchart