**CHAPTER 1**

**INTRODUCTION**

* 1. **Introduction and Problem Statement**

The rise of technology and the internet has had a tremendous impact on the way people in Thailand conduct their daily activities, including buying and selling products. E-commerce platforms such as Shopee and Lazada have revolutionized the way people shop, and this has been reflected in the projected growth of the E-commerce market in Thailand by 15-20% per year, according to a survey. These platforms have become central applications in Thailand, providing a convenient way for consumers to access a vast array of products, with the added feature of allowing customers to read reviews of previously purchased products. This feature has been found to be a significant reason why Thai people choose to buy products online through E-commerce platforms, with a high percentage of respondents indicating its impact on their purchasing decisions, as found in the Digital Stat report.

Among all the groups of products sold on these platforms, IT gadgets are particularly popular and highly demanded in the marketing industry. However, these IT products can vary significantly in their usage characteristics, and they have specific quality indicators or descriptions that can pose a challenge for consumers. This variation makes it difficult for consumers to make informed purchase decisions based on the product descriptions provided. As a result, consumers who read reviews may need to spend a lot of time reading a large number of reviews to find good ones that can help them make purchase decisions, especially for consumers who are not knowledgeable in the IT product group. In addition, they may also receive unreliable reviews, which further results in buyers losing confidence in purchasing those products.

One way to solve the above problem is to use a machine learning model to filter out useful product reviews and reduce the time for consumers to read a large number of reviews. However, to create such a model, a training dataset that addresses this problem is required. Unfortunately, in Thailand, there is a lack of data that defines useful product reviews, especially for product reviews in the IT product group that can help consumers. This presents a significant challenge for researchers who are trying to develop a machine learning model that can classify useful product reviews in the Thai language.

Moreover, although there are research studies in Thailand that address the problem of filtering product reviews, they often use data that is already popular, making it difficult to generalize the findings. Even when the researchers collect their own data, they do not provide a detailed explanation of how they define or answer the data, or how they collect the data, especially for the IT product review group. Therefore, it is crucial to have a well-defined and comprehensive training dataset that can help researchers develop and evaluate a machine learning model that can effectively classify useful product reviews in the Thai language.

In conclusion, while E-commerce platforms have revolutionized the way people shop in Thailand, the problem of finding trustworthy product reviews remains a significant challenge, especially for highly demanded IT gadgets. Using a machine learning model can help filter out useful product reviews and reduce the time for consumers to read a large number of reviews. However, creating such a model requires a comprehensive training dataset that addresses the problem, particularly for the IT product group in the Thai language. Researchers need to collect and define a well-defined and comprehensive dataset to develop an effective machine learning model to classify useful product reviews in the Thai language.

Top of Form

* 1. **Objectives**

Develop a machine learning model that can effectively filter out useful product reviews in the Thai language for IT products

Collect and create a comprehensive dataset of useful product reviews in Thai for the IT product group

* 1. **Significance of the Research**
  2. **Scope of Study**

**CHAPTER 2**

**LITERATURE REVIEW**

Online product reviews have become an essential source of information for online consumers, particularly for those who shop on e-commerce platforms. However, with the increasing number of reviews available, it becomes increasingly challenging for consumers to determine the most helpful ones. To address this issue, researchers have explored different approaches to determine the helpfulness of online product reviews. This literature review aims to examine the different methods used to determine review helpfulness and evaluate their potential for improving online consumer decision-making.

**2.1 Determining the helpfulness of online product reviews**

Determining the helpfulness of online product reviews has become a crucial factor in the decision-making process of consumers. With the rise of e-commerce, the importance of online product reviews has increased, as they provide valuable insights for potential buyers. Researchers have identified various factors that contribute to the perceived helpfulness of online product reviews.

Li et al. (2013) define review helpfulness as the degree to which reviews facilitate purchase decisions. They identify three dimensions of helpfulness: source credibility, content diagnosticity, and vicarious expression. Source credibility refers to the perceived trustworthiness of the review author, while content diagnosticity relates to the review's ability to provide useful information for potential buyers. Vicarious expression is the degree to which the review reflects the buyer's own experiences or expectations.

Mudambi and Schuff (2010) explore the impact of review depth and extremity on review helpfulness, taking into account the product type being evaluated. They propose several criteria to determine helpful reviews, such as review length, information diagnosticity, and review extremity. Review extremity refers to the degree of positive or negative sentiment expressed in the review.

Siering and Muntermann (2013) present a research model that considers review extremity, depth, and product type as factors influencing review helpfulness. They expand the model to include product quality, review sentiment, and uncertainty, and examine their impact moderated by product category. Product quality refers to the product's reliability and durability, while review sentiment relates to the emotional tone conveyed in the review.

Cao et al. (2011) study the determinants of helpfulness for online user reviews of software programs. They find that semantic characteristics, such as opinions, level of detail, and technical terms, have the greatest impact on review helpfulness. They suggest designing effective mechanisms for helpfulness voting to encourage users to write more meaningful comments. Reviews with more extreme opinions are more likely to receive more helpfulness votes, according to the study's findings.

In conclusion, determining the helpfulness of online product reviews is crucial in the decision-making process of consumers. Researchers have identified various criteria, including product reliability, diagnostic, sentiment, quality, and semantic characteristics, that contribute to review helpfulness. Online retailers can use text analysis and machine learning classifiers to identify helpful reviews and promote them to potential buyers. Effective helpfulness voting mechanisms can also encourage users to write more meaningful comments, thus improving the overall quality of online product reviews.

**2.2 Implementation of a recommendation system in product reviews.**

Online shopping has become the preferred method for many consumers, with product reviews playing a critical role in purchase decision-making. However, the sheer volume of reviews for even moderately popular products can make it challenging for consumers to find the most helpful ones. This is where a recommendation system in product reviews can be beneficial. By using machine learning algorithms to analyze review content and provide personalized recommendations, a recommendation system can help consumers make more informed purchasing decisions, while also improving customer satisfaction and loyalty for businesses.

This study focuses on developing a machine learning-based model to predict the helpfulness of consumer reviews by evaluating different determinants such as readability, polarity, subjectivity, entropy, and average review rating. The research aims to mitigate the Matthew effect by automatically assigning helpfulness values to an initial review as soon as it is posted on the website. The study found that ensemble learning techniques, specifically gradient boosting, outperformed linear regression techniques in terms of performance metrics such as mean squared error. The research also revealed that for experience products, the star rating was a more important parameter than for search products. However, for both types of products, the readability, entropy, and sentiment parameters had a similar effect. The limitations of the study include the exclusion of non-English words and the analysis of data from only one e-commerce website, which may limit generalizability. Future research can consider more extensive datasets and examine the effectiveness of the model on other e-commerce websites.