

University of Applied Sciences

Exposé (05/04/2023)

Name: Henny Purwadi.

Supervisor: Prof. Dr. Frank Passing

Study programme: Bsc.Data Science

Enrolment number: 32009177

Working title: A Comparative Sentiment Analysis of Digital and Robot Pet Companions in Various Locations

Overall aim: The purpose of this thesis is to find out whether location differences affect human sentiment towards digital and robot pet companions or not, by using machine learning model for sentiment analysis and testing hypothesis on unlabeled data collected from various locations.

Objectives: The objective of this study is to provide insights about how geography and location may affect or not affect humans' sentiments regarding digital and robot pet companions.

Abstract

This study is to find out whether geographic locations have an impact on emotion towards robotic and digital pet companions, using machine learning models and hypothesis testing. This investigation will show and assess different machine learning models for sentiment analysis on labeled datasets and pick the model with the best performance. This exploration will collect unlabeled data about digital robot companions from various regions, pre-process the information, and apply a chosen machine learning model to figure out sentiments in the unlabeled dataset. Then use chi-square to evaluate the connection between area and sentiment towards robot companions and find out whether there is a critical distinction in the sentiment between various locations. There will be discussion about the result, limitations, and its suggestions for understanding human-machine relationships in various zones. The investigation will give potential results for future research.

Keywords: Sentiment Analysis, Machine Learning, Robot Companion, Social Robot, Artificial Intelligence, Machine Learning, Natural Language Processing (NLP), Logistic Regression, Naive Bayes Classification, Linear SVC, and TF-IDF Vectorization, GPT-3 zero-shot classifier, Transformers, GPT-2, BertForSequenceClassification.

1 Introduction

1.1 Background and Motivation

If anyone have a dream that one day artificial intelligence (AI) and humans will unite, then one of the most effective ways to achieve this goal is to start gradually developing emotional bonds between humans and AI, by having AI as digital pet companions or virtual human assistants for humans. Creating an emotional connection between humans and AI, whether through digital pet companions or virtual human helpers, is potentially an efficient approach to moving towards the goal of achieving the human-machine unification.

These days, it is not uncommon for children and adults to have their own personal digital robot pet as a companion. This new generation grew up with artificial intelligence as part of the family, so they are already comfortable with it.

It is possible that it will be easier for people to embrace AI and incorporate it into their lives if humans and AI have been able to build trust and familiarity with one another. Additionally, starting with introducing them to AI in the form of digital pet companions can be a valuable way to shape their perception of AI and potentially pave the way for wider acceptance of AI in the future.

The researcher was motivated to carry out this research by the story of Jibo, Emo, Vector etc., social robots designed to be friends for humans. Their owners' feelings of loss, hopelessness, and disappointment as they contemplate the possibility of parting ways with their beloved robot companions.

This sentiment suggests that people can form deep bonds with their robotic pets, just as they do with other members of their families. As artificial intelligence (AI) and robots advance, it is important to think about how they will affect people emotionally.

The issue is whether people around the world share the same emotional attachment, such as joy at the presence of robot pets and social robots, and the same sadness in their absence? This study is designed to answer this question.

1.2 Research Questions and Objective

Research Questions

1. Will the majority sentiments about digital and robot pet companions be positive, like joy & love, or will be negative, like anger & fear?
2. Does location/ region influence people's sentiments towards digital and robot pet companions?

Research Objective

The main objective of this study is to understand the relationship between region and sentiment towards digital and robot pet companions.

1.3 Value of Study and Target audience

The value of this study is that it has the potential to give understanding regarding how component like area can affect people sentiments of robot pet and digital companions.

Target audience

The discoveries from this research might help organizations and associations that wish to advance digital robotic pet companions in a variety of fields. Researchers and academics interested in human-machine collaboration and the social components that influence this relationship may also find this study useful.

1.5 Structure of the Document

The structure of this research will be as follows: Chapter 1 will explain the research's background & motivation, research questions and goals. Chapter 2 will discuss the theoretical background and review of previous research related to human-machine emotional relationships and machine learning usage in sentiment analysis. Chapter 3 will cover the research methodology, including unlabeled data collection, data pre-processing, and ethical considerations. Chapter 4 will cover the results and findings. Chapter 5 will discuss interpretation of results and limitations of the study. Chapter 6 will provide a summary of the study's findings and recommendations for future research.

1.6 Scope and Limitations

Machine learning used to identify the sentiment of English-language unlabeled dataset gathered from different areas without separating by gender or age. There are drawbacks like bias in sentiment categorization algorithms based on machine learning. Other than that, people are more affected by negative events instead of by positive ones. For millions of years, our brains have evolved to respond negatively to threats, as proposed by psychologist Rick Hanson. (Hanson, 2019).

2. Literature Review

2.1 Overview of Relevant Previous Studies

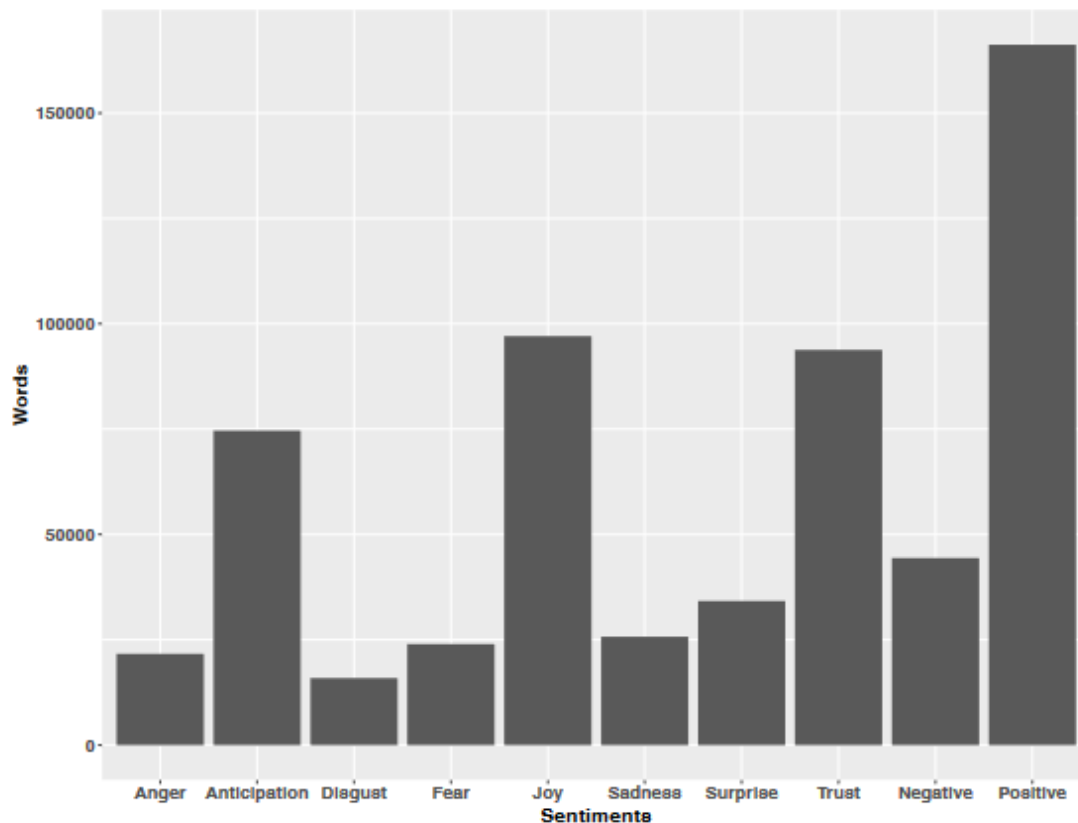
A summary of prior research on human-machine emotional interaction and sentiment analysis from "Why Do We Turn to Virtual Companions? A Text Mining Analysis of Replika Reviews" (Siemon, Strohmann, Khosrawi-Rad, Elshan, de Vreede, & Meyer, 2022):

According to Siemon et al. (2022), conversational agents are intelligent systems that allow users to interact with them using natural language. Using social chatbots aims to provide the user with a natural and human-like interaction. This type of interaction is especially desired for longer term interactions as it helps to build relationships between the user and the conversational agents.

The previous study's goal was to discover the topics and emotions that users experienced when communicating with Replika, a digital companion, based on 119,831 reviews of Replika collected by previous researchers from the Google Play Store and then subjected to sentiment analysis and topic modeling.

The results of the sentiment analysis show that most users really like the Replika application, and most users feel joy, happiness, and a sense of well-being.

Figure 1. Results of the sentiment analysis



Source: Siemon et al. ,2022, p. 5

Trust and **joy** were two of **the positive** feelings that consumers reported feeling towards the virtual companion, besides **anger**, **sadness**, and **fear** as negative feelings which may reflect a general apprehension about new forms of artificial intelligence.

Table 1. Results of the topic modeling

Topic name	Topic 1 – Excitement about AI	Topic 2 - Companionship	Topic 3 – Well-Being and Support	Topic 4 – Enjoyment
Relevant words	AI Good Fun Pretty Cool Interesting Conversation	Like Real Talking Person Someone Friend Nice	Asked Great Feel Good Helpful Better Talk	Amazing App Love Good Awesome Fun AI

Source: Siemon et al. ,2022, p. 5

The results of the topic modeling analysis show that the most important thing that users talk about is that they feel better and enjoy using the application. Siemon et al. (2022, p. 5)

This research found that most people use Replika because it makes them feel good, overcomes loneliness, and provides emotional support. However, this study reveals negative reviews regarding virtual friends due to expectations that are too high. The study also indicates that further research is needed to look into possible negative effects and ethical issues in relationships with virtual friends, and how the use of this technology affects relationships and human well-being in the long term.

2.2. Terminology and Definitions

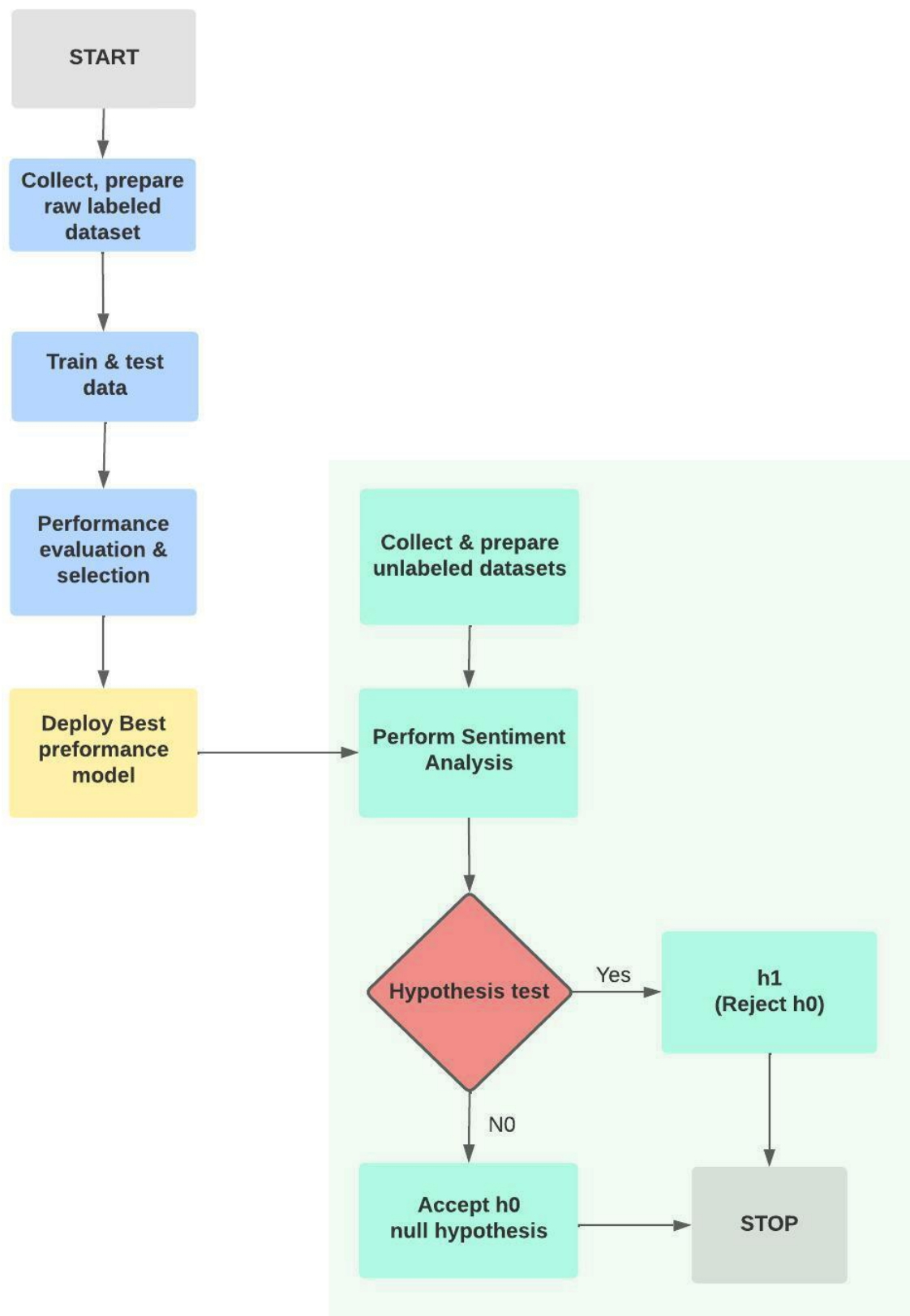
Terminology and definitions related to widely used machine learning algorithms for sentiment analysis, such as Naive Bayes, Logistic Regression, Linear SVC, Transformer, GPT, Bert, will be explained in this chapter, as well as formulas behind each machine Learning algorithms used in this study.

3 Methodology

The methodology for this study will be a combination of approaches, between:

1. An experimental approach for comparing several model performance, and model selection based on the best performance.
2. A Quantitative approach for data collection, sentiment analysis, and conduct hypothesis testing.

Figure 2. Sentiment analysis workflow



Source: Own source

1. Train and Evaluate performance of Machine Learning Models on labeled datasets.

Choose a machine learning model in Python for sentiment analysis, including Naïve Bayes, Logistic Regression, LinearSVC, and compare their performance.

2. Compare Performance of the different machine learning models.

3. Collect Unlabeled Data about digital pet companions and robots from various locations, such as in America, Europe, Asia, Australia, and Afrika from 2014 to 2023, using Python's library using Python's library.

4. Ethical Considerations, including responsible and respectful data collection and use.

Ethical considerations for this research will ensure that data is collected and used in accordance with relevant ethical guidelines and regulations. s will be collected from publicly available sources that are open to the public and will not contain any personal or sensitive information. Because this study is based on published data, informed consent is not required.

5. Pre-process and Vectorize Collected Data by cleaning and filtering the data s with Python's library re and nltk and preparing the data for analysis. Then tokenize and lemmatize the data, and vectorize using the tf-idf vectorizer.

6. Predict Sentiment on unlabeled s using the trained machine learning models, classifying them as joy, sadness, anger, love, fear, or surprise based on the language used in the s in each region separately.

7. Hypothesis Testing

Next, we use the Chi-square Hypothesis test to determine whether there is a relationship between location and sentiment towards companion robots, and to determine whether there are significant differences in the distribution of sentiments between different regions.

Null hypothesis h0:

Joy, Sadness, Surprise, Anger, Love, Fear Sentiments are INDEPENDENT upon different continents.

Alternative hypothesis h1:

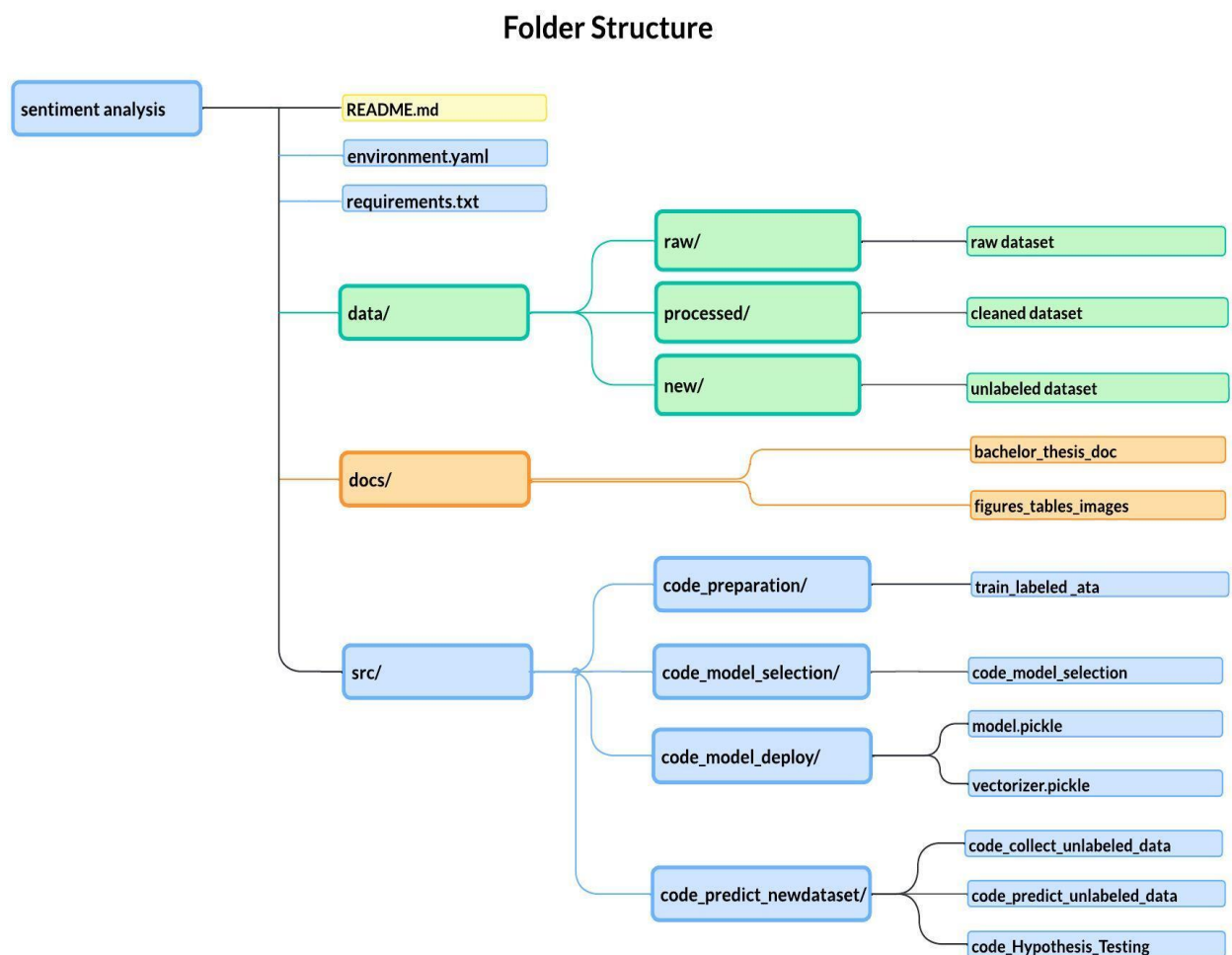
Joy, Sadness, Surprise, Anger, Love, Fear Sentiments are DEPENDENT upon different continents.

This study can conclude that there is a dependence between location and sentiment on human-machine relations if the results of the chi-square test show a significant difference in the distribution of sentiment between regions. It means that people's sentiment towards companion robots is independent depending on/influenced by their geographical location.

However, if the chi-square test reveals that there is no significant difference in the distribution of sentiment across regions, then we cannot reject the null hypothesis, and forced to draw the conclusion that there is no dependence between location and sentiment regarding human-machine relations.

This means that people's sentiment towards companion robots is independent, not influenced by their geographical location.

Figure 3. Folder Structure



Source: Own source

8.Expected Outcome:

1. The expected outcome is that logistic regression method will outperform other methods like Naive Bayes and Linear SVC.
2. Initial beliefs are that the sentiment analysis outcome would be consistent with earlier research in chapter two, which indicated more positive emotions such as joy and love than negative emotions such as anger and fear.
3. The researcher is unsure whether there will be dependencies between locations and human feelings about digital companions.

4. Results and Findings

4.1. Labeled Data Preparation

4.2. Performance Evaluation and Model Selection

4.3. Unlabeled Data Collection

4.4. Sentiment Analysis Result on Unlabeled Data and Location Comparison

4.5. Hypothesis Testing Result

5. Discussion:

5.1. Result Interpretation

5.2. Discussion of Limitations

6 Conclusion

6.1. Summary

6.2. Recommendations for Future Research

Preliminary reading list:

Siemon, Strohmann, Khosrawi-Rad, Elshan, de Vreede, & Meyer. (2022, July 11). Why Do We Turn to Virtual Companions? A Text Mining Analysis of Replika Reviews. Retrieved from https://aisel.aisnet.org/amcis2022/sig_hci/sig_hci/10/