# Aspect Based Sentiment Analysis of Finnish Neighborhoods: Insights from Suomi24

**Laleh Davoodi** and **Anssi Öörni** and **Ville Harkke**
Faculty of Social Sciences, Business and Economics, and Law, Åbo Akademi University
first.last@abo.fi

## Abstract

This study presents an approach to Aspect-Based Sentiment Analysis (ABSA) using Natural Language Processing (NLP) techniques to explore public sentiment across 12 suburban neighborhoods in Finland. We employed and compared a range of machine learning models for sentiment classification, with the RoBERTa model emerging as the best performer. Using RoBERTa, we conducted a comprehensive sentiment analysis(SA) on a manually annotated dataset and a predicted dataset comprising 32,183 data points to investigate sentiment trends over time in these areas. The results provide insights into fluctuations in public sentiment, highlighting both the robustness of the RoBERTa model and significant shifts in sentiment for specific neighborhoods over time. This research contributes to a deeper understanding of neighborhood sentiment dynamics in Finland, with potential implications for social research and urban development.

## 1 Introduction

Understanding public sentiment towards specific neighborhoods is crucial for urban planners, sociologists, and policymakers, as it offers insights into the social fabric of communities and the challenges they face. With the advent of social media and online forums, the internet has become a rich source of data reflecting public opinion. Suomi24[1], the largest Finnish online forum, is an ideal platform for studying the sentiments of neighborhoods through user-generated content. Hence, we used Suomi24 as our data source driven by the following factors. First, social media has become a popular platform for expressing opinions, making it valuable for SA on topics such as urban issues, climate change, and healthcare (Huang et al., 2021; Chen and Wei, 2023). Second, online forums provide an efficient way to share information and generate discussions, offering a wealth of informative comments that can be analyzed for sentiment (Chen and Wei, 2023). Building on this, (Lindén et al., 2023) utilized the Suomi24 data to create a manually annotated dataset[2] for the SA task, further highlighting the platform's potential for research in Finnish SA.

The field of urban design and planning increasingly utilizes SA to evaluate urban environments through crowdsourced data (Tas and Sanatani, 2023). Platforms like Point of Interest (POI) databases, social media, and citizen engagement tools have proven valuable for large-scale urban assessments (Martí et al., 2019; Huang and Gartner, 2016; Tas and Sanatani, 2023). There are several techniques in Literature using NLP to extract sentiment from geo-located textual data.

However, most models focus on identifying general positive or negative sentiments within a text, with limited efforts made to train models that can detect specific aspects and their associated sentiments (Tas and Sanatani, 2023). Therefore, to expand the limited research on ABSA in Finnish suburbs, we seek to explore the following research question (RQ): *How do sentiments toward suburbs in Finland, expressed in public online discussions, vary over time, and how effectively can ABSA models capture and predict these sentiments?*

To address this research objective, we investigate public sentiment toward 12 different areas in Finland by utilizing a manually annotated Finnish-language dataset for the ABSA task. For this purpose, each neighborhood is treated as a distinct 'aspect,' allowing us to classify sentiments expressed toward these locations rather than identifying specific reasons for likes or dislikes associated with each neighborhood. This approach provides an overview of the sentiment trends for each neighborhood over time, using

[1] https://www.suomi24.fi/

[2] https://clarino.uib.no/comedi/editor/lb-2023012701



*Proceedings of the 9th International Workshop on Computational Linguistics for Uralic Languages*, pages 1–11
November 28-29, 2024 ©2024 Association for Computational Linguistics

the neighborhood names as predefined aspects for classification. Afterwards, we applied a combination of classical machine learning models such as Support Vector Machines (SVM) and Naive Bayes, alongside modern deep learning models like BERT[3](TurkuNLP/bert-base-finnish-cased-v1)(Virtanen et al., 2019) and RoBERTa (Finnish-NLP/roberta-large-finnish-v2)[4], to analyze sentiment towards these neighborhoods. Our results show that RoBERTa outperforms other models in accurately predicting sentiment across these neighborhoods.

Further, we expanded our analysis by using RoBERTa to predict sentiment on a larger dataset of approximately 32,000 data points for the suburbs of interest in Finland. This allowed us to track trends and shifts in sentiment over time, providing valuable insights into how public perceptions of different neighborhoods evolve.

Our findings contribute to the growing body of research on SA and urban studies, particularly in the Finnish context. The results have practical implications for regional development and policy formulation, as understanding public sentiment is key to fostering better community relations and addressing regional disparities.

## 2 Sentiment analysis and social media data

In the age of digitalization, the growth of social media has significantly transformed the global flow of information and the organization of social demands. Platforms such as Twitter, Facebook, and Instagram have emerged as vast repositories of real-time opinions from diverse groups (Troya et al., 2021). The vast volume of online-generated data makes manual analysis challenging and hinders the ability to identify trends in a timely manner. As a result, NLP techniques, particularly SA, have become widely used in recent literature for analyzing social media content and user feedback (Benrouba and Boudour, 2023; Xu et al., 2022). Sentiment analysis/classification is a specific type of text categorization, where the classification is based on the author's expressed attitude by understanding of the document and focusing on how the sentiment is conveyed throughout the text (Meena and Prabhakar, 2007).

SA is usually performed at three levels as follows: document level, sentence level, and Aspect/category level. The objective of document-level sentiment analysis is to assess the overall sentiment of the entire document. Sentence-level sentiment analysis evaluates the overall sentiment of a sentence as a single entity and predicting the opinion expressed within that sentence (Do et al., 2019). Whereas ABSA goes deeper by identifying specific opinion targets (aspects) within the text and determining the sentiment associated with each target (Karimi et al., 2021).

SA techniques are typically divided into supervised, unsupervised, and semi-supervised learning approaches. The supervised method applies labeled data for training and evaluation. The lexicon-based approach is an example of an unsupervised method, where the sentiment of a document is determined by assessing the semantic polarity of individual words by utilizing a pre-existing dictionary that assigns sentiment polarities to words or phrases. The overall sentiment of the text is then determined based on the predominant polarity of the words. (Taboada et al., 2011).

Traditional machine learning techniques, including SVM (Ahmad et al., 2018; Sharma and Sabharwal, 2019) and Naive Bayes (Hariguna and Rachmawati, 2019; Wongkar and Angdresey, 2019), have been extensively utilized in various SA applications. Traditional machine learning methods often rely on Bag-of-Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF) for feature extraction. However, these approaches tend to generate sparse, high-dimensional feature vectors and have limited ability to capture the word order, syntactic structure, and semantic meaning of a sentence (Kokab et al., 2022).

Recent advancements in deep learning models have surpassed the performance of traditional methods by addressing the limitations of traditional machine learning methods. Bidirectional Encoder Representations from Transformers (BERT) (Devlin, 2018) is a pre-trained model that utilizes a self-attention mechanism (Vaswani, 2017) to weigh the importance of each word in a sentence, and captures complex word dependencies regardless of their position. Unlike traditional models, BERT processes text bidirectionally to understand the full context, enabling deeper comprehension of syntactic and semantic relationships.

RoBERTa (Robustly Optimized BERT Pretraining Approach)(Liu, 2019) is a variant of the BERT

[3]https://huggingface.co/TurkuNLP/bert-base-finnish-cased-v1

[4]https://huggingface.co/Finnish-NLP/roberta-large-finn

model, that improves BERT's performance by modifying its training process and showing better performance on natural language understanding(NLU) tasks (Cortiz, 2022). RoBERTa removes BERT's next sentence prediction task, increases the training data, and extends the training duration with larger mini-batches and more diverse data sources (Liu, 2019).

#### 2.0.1 Related Work

While numerous studies have explored SA of social media across various domains, such as tourism (Mehra, 2023; Ali et al., 2021) and e-commerce (Vanaja and Belwal, 2018; Davoodi and Mezei, 2022), relatively few have focused on SA in the context of urban environments. As an example, Saeidi et al. (2016) focus on identifying sentiments toward specific aspects of multiple entities within the same text. This work extends traditional ABSA by allowing multiple entities and aspects in a single document and presents the SentiHood dataset, which contains user discussions about urban neighborhoods from a question-answering platform. The dataset includes annotated sentences where various aspects (e.g., safety, price, or transit) of different neighborhoods are discussed. The paper also provides strong baselines using logistic regression and LSTM models and analyzes their performance on this new, more complex SA task.

Another relevant study, Rui (2023) uses machine learning models to reveal the spatial heterogeneity in sentiment, showing a medium-high-low trend from the city center to suburban areas. Factors such as metro route density and walkability were positively correlated with sentiment in formal settlements. The findings provide urban planning recommendations for promoting positive sentiments and achieving sustainable development goals.

Wang (2023) presents a methodology for incorporating public perception into urban planning evaluation using SA. By analyzing comments from social media platforms such as Weibo and Xiaohongshu, the authors develop a public perception-oriented evaluation model. The approach captures sentiment across aspects like urban landscape, ecological environment, and living quality. The results demonstrate that H-CAN outperforms existing models, achieving 84.2% accuracy on the Yelp 2015 dataset. The study by Tas and Sanatani (2023) focuses on evaluating urban environments using geo-located data, particularly through aspect-based analysis. The authors assess different aspects

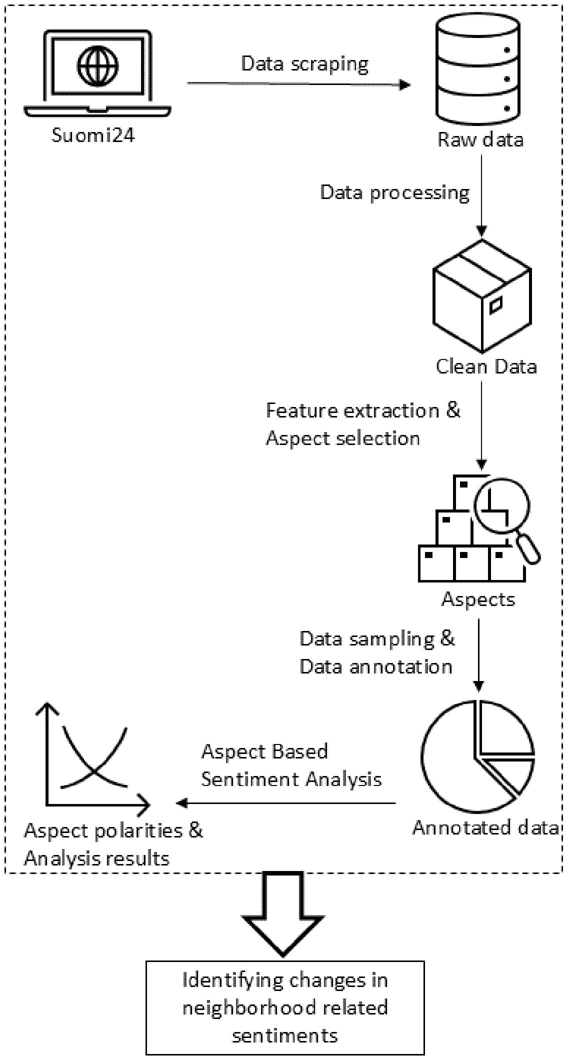



Figure 1: The stages of the research process

of urban living conditions by analyzing various sources, such as user feedback or environmental data. They annotated a dataset consisting of 2,500 crowdsourced reviews of public parks and trained a BERT model with a Local Context Focus (LCF). Their model demonstrates a substantial improvement in prediction accuracy for urban reviews.

## 3 Methodology

In this section, we outline our methodology for conducting ABSA using user-generated content. The objective is to analyze and understand people's sentiments toward 12 suburban neighborhoods in Finland. By identifying sentiment changes over time for these areas, we aim to uncover patterns related to public perception and community sentiment changes.

As depicted in Figure 1, our process begins with scraping data from Suomi24, followed by data processing. In the feature extraction and aspect selection phase, we identify and tag the specific areas of interest within the text. The data is then sampled and annotated with corresponding sentiment labels, preparing it for ABSA. The output consists of sentiment polarities for each area, which we analyze to detect trends and shifts in public sentiment over time. This methodology helps in understanding how residents and visitors perceive these neighborhoods and track any significant changes in sentiment.

### 3.1 Data collection and processing

To build our Finnish-language dataset, we collected data from the Suomi24 corpus, available through the Language Bank of Finland[5]. This corpus includes over 84 million messages posted on Suomi24, the largest Finnish online discussion forum (Lagus et al., 2016).

Finnish, being an agglutinative language with complex morphology, presents some challenges in text analysis (Vilkuna, 1989). However, the lemmatized corpus simplifies this by presenting words in their base forms, allowing for easier identification of lexical morphemes. For this study, we focused on twelve suburbs that have undergone systematic urban planning. To collect relevant messages, we used the suburb names as lemmas, retrieving all posts mentioning at least one of these areas from 2001 to 2017. This yielded a dataset of 36491 relevant messages. These areas, in descending order according to the number of messages are Kontula, Varissuo, Leppävaara, Myllypuro, Matinkylä, Maunula, Uittamo, Suikkila, Harittu, Runosmäki, Soukka, and Pihlajamäki. After that, we randomly selected 3183 messages to be annotated manually by 2 native Finnish speakers.

In order to perform ABSA, we opted for a manual sentiment annotation process due to the lack of a domain-specific dataset for this task in the Finnish language. To the best of our knowledge, there is no publicly available dataset in Finnish specifically for the ABSA task. In his Master's thesis, (Hellström, 2022) presents a dataset collected from public reviews on Verkkokauppa.com, focusing on laptops and tablets. The dataset consists of 1,673 sentences, which were manually labeled by the author following the guidelines established by SemEval 2014 for SA tasks (Hellström, 2022). In contrast, our dataset is tailored to analyze sentiment across different neighborhoods in Finland, specifically for the ABSA task, with careful resolution of any disagreements during annotation to ensure domain-specific accuracy. In addition to the ABSA dataset, we developed a lexicon-based dictionary by compiling 33742 unique words from our manually annotated dataset assigned with positive and negative scores by a native Finnish speaker annotator.

The aspects are specified as the 12 interest areas mentioned above. Three possible sentiment labels were assigned to each aspect-message pair: positive, neutral, and negative. The two annotators individually assigned sentiment values to each aspect. Every disagreement was marked and discussed to reach the final agreement between the annotators. All the messages of inadequate quality (e.g., the aspect mentioned in the text does not refer to an area) were excluded from further analysis. The final annotated dataset contains 3183 data points.

After finalizing the data, we calculated the most frequent words to find the most frequent Topics discussed in the text. For instance, the most discussed topic in the Varissuo area is immigration, with terms like "maahanmuuttaja" (immigrant) appearing 437 times, alongside related words such as "pakolainen" (refugee) with 79 occurrences and "muslimi" (Muslim) mentioned 154 times. Other terms related to demographics, such as "väestö" (population) and "kantaväestö" (indigenous population), also appear frequently, highlighting the ongoing discussions about immigration and its impact on the area. Another prominent topic is education, with "koulu" (school) being mentioned 356 times. Public transport is also commonly discussed, with terms like "bussi" (bus) at 309 occurrences and "joukkoliikenne" (public transport) appearing 155 times, indicating the importance of infrastructure in the area. Discussions about housing, crime, and services also frequently appear, reflecting various community concerns.

Finally, we performed 4 steps to clean the text: 1) transforming the text into lowercase, 2) removing all non-alpha characters, 3) removing HTML tags and URLs, 4) and eventually tagging specific aspects of interest using custom markers like '<TAG>aspect</TAG>'. Afterwards, to address the imbalance in positive sentiment data, we applied a data augmentation technique using back-translation. Back-translation has emerged as a popular data augmentation technique, where a reverse

[5] https://www.kielipankki.fi

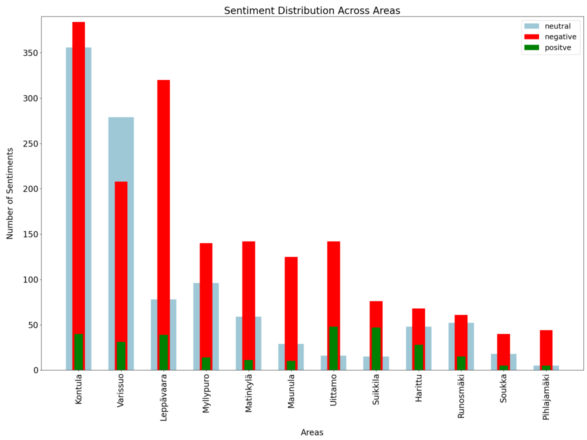



Figure 2: The distribution of Areas in Annotated Dataset

translation system is used to translate monolingual text from the target language back to the source language (Pham et al., 2023). We translated the positive text from Finnish to German and then back to Finnish, generating additional data with subtle variations while maintaining the original sentiment. This augmented data was then combined with the original dataset to achieve a more balanced distribution of positive sentiments for model training. Finally, after translating and tagging the text, we performed downsampling to address class imbalances in the sentiment labels. Since the neutral class was overrepresented, we downsampled the neutral sentiment examples to match the number of negative and positive sentiment examples. This resulted in a more balanced dataset, improving the performance of the model in training and reducing bias towards the majority class.

After performing the data processing, the dataset consists of a total of 2,688 instances. The distribution of sentiment labels within the dataset is as follows in order: Negative Sentiment (1,051 instances), Neutral Sentiment (1,051 instances), and Positive Sentiment (586 instances). The distribution of different areas is presented in Figure 2.

## 3.2 Machine learning models for sentiment classification

There have been several traditional machine learning (Wagner et al., 2014; Tang et al., 2019) and deep machine learning (Talaat, 2023; Tas and Sanatani, 2023) models that are proposed to solve the ABSA tasks. Based on the results of various studies, deep learning methods tend to show more adequate performance at defining the complex nonlinear correlations between the features and the sentiment polarity. In particular, neural network-based methods have outperformed other methods for ABSA tasks because these methods are trained end-to-end and can comprehend significant features automatically (Jiang et al., 2019). Considering all these observations, in this study, we have chosen models in literature for social media review as: (i) the most commonly used traditional classification models for comparison purposes (Naïve Bayes, Support Vector Machines), (ii) deep learning models that have shown the best performance in prior studies (BERT, and RoBERTa).

### 3.2.1 Model building

We tested different varieties of text feature extraction techniques and machine learning algorithms to specify the best-performing model. In order to transform the text into a set of understandable features for the traditional classifiers, we utilized BOW and TF-IDF. BOW converts text data into a representation based on the frequency of word occurrences within a document, ignoring the order and syntactic structure of the words (Yan et al., 2020). TF-IDF is a statistic approach that defines the weight for each term (or word) in each document based on frequency and informativeness (Soucy and Mineau, 2005).

For the Naïve Bayes and SVM classifiers, we employed BOW and TF-IDF to generate the numeric features, while for BERT and RoBERTa, we used transformer tokenizers. For the non-deep learning models, we divided the dataset into the train (80% of the data), validation (12%), and test (8%) sets using stratified sampling. The optimal cost parameters on the validation set were identified: (i) by adjusting the n-gram vectorizer parameters and regularization parameters for the linear SVM model, and (ii) by n-gram vectorizer parameters and the $\alpha$ value for Naïve Bayes. After defining the optimal hyperparameters for each model, we selected the best-performing model to evaluate the number of misclassifications on the test set.

In the case of transformer-based models, we used cross-validation to have a realistic estimation of our models' performance. We split the dataset into training and test sets by using four-fold cross-validation. We utilized transformer tokenizers: AutoTokenizer for BERT, and RobertaTokenizer for RoBERTa. In both cases, we added a dropout layer to control overfitting in addition to the baseline. When training RoBERTa, the model includes the main model with 12 layers and 768 hidden dimen-

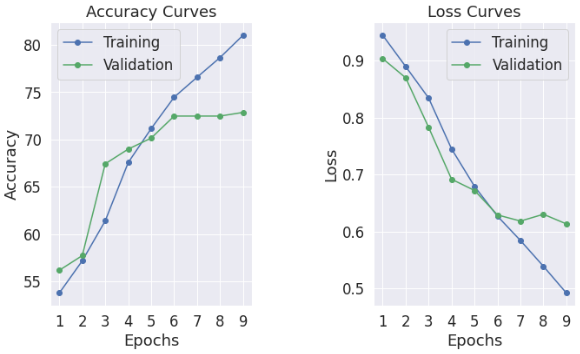



Figure 3: The performance of the best model - RoBERTa

sions. For both BERT and RoBERTa, the optimizer, loss function, and performance metric were defined as Adam, categorical cross-entropy, and accuracy. For the RoBERTa model the initial learning rate was 3e-5, and the batch size was defined as 8.

### 3.3 Results

In this section, we present the results of SA, with a detailed comparison of both the performance of different models and the sentiment trends across these areas/aspects considered in the study.

#### 3.3.1 Sentiment classification performance

To quantify the performance of each model for the ABSA task, we utilized weighted F1 score and accuracy. The results are shown in Table 1. Based on the results, RoBERTa slightly outperforms BERT in terms of the F1 score (0.75 vs. 0.74). Since F1 score is crucial for balancing precision and recall, RoBERTa is the best model for this task. Even though RoBERTa has a lower training accuracy compared to other methods (e.g., Naive Bayes or SVM), it generalizes better on the test set.

To build the best model for ABSA, a fine-tuned version of RoBERTa was utilized with a fixed dataset split to ensure robust evaluation. Prior model training, to train the best-performing model, we also utilized a lexicon-based dataset to shorten long texts. For this purpose, we filtered neutral sentences by removing sentences with an overall neutral lexicon score. The overall score was calculated by summing the lexicon scores of all the words in a sentence. If the sum was 0, the sentence was considered neutral; if the score was greater than 0, the sentence was classified as positive, and if it was less than zero, the sentence was classified as negative. Afterwards, the dataset was preprocessed to handle instances where multiple areas (aspects) were mentioned within a single row. For rows that referenced more than one area, the sentence was split into separate rows, with each row corresponding to a specific area mentioned. After this preprocessing step, the final training dataset comprised 3,678 rows, ensuring that each entry focuses on one area for a more accurate sentiment and aspect-based analysis. Finally, the dataset was divided into fixed splits with 3,318 samples for training, 111 samples reserved for testing, and 258 samples for validation. As shown in Figure 3, the accuracy and loss curves demonstrate effective learning, with training accuracy steadily increasing while validation accuracy stabilizing after epoch 5, suggesting that the model generalizes well without significant overfitting. The number of misclassifications on the test set was 31 out of 111 datapoints. The fine-tuned RoBERTa model was subsequently employed to predict sentiment trends across the rest of the 32,183 data points.

Regarding the impact of aspect tagging, <TAG>aspect</TAG>, we evaluated model performance both with and without tagging. The results indicate that tagging aspects in the text improve both training and test accuracy. For instance, the tagged version of the best model,RoBERTa, achieved a test accuracy of 72.07%, compared to 63.37% for the untagged version. This suggests that tagging provides additional contextual cues, helping the model more effectively identify and associate sentiments with specific aspects. Additionally, Finnish morphology presents unique challenges, as aspect names can appear in multiple inflected forms (e.g., 'Uittamo' could appear as 'Uittamolla', 'Uittamolta', or 'Uittamolle'). Without tagging, the model may struggle to recognize these variations as referring to the same aspect, potentially impacting its ability to learn consistent sentiment associations. Tagging standardizes the aspect representation within the text, helping the model to recognize it regardless of morphological variation, which is particularly beneficial in Finnish.

#### 3.3.2 Trend analysis

In order to calculate sentiment trends for each area, we weighted sentiments by assigning values of -2 for negative, 0 for neutral, and +2 for positive sentiments, effectively minimizing the impact of neutral sentiments. Finally, for each area, we summed the weighted sentiment scores by year and normalized the final sum by dividing it by the total

| Method | Training set accuracy | Test set accuracy | F1 on test set |
|---|---|---|---|
| Naive Bayes (TFIDF) | 0.95 | 0.67 | 0.65 |
| Naive Bayes (BOW) | 0.94 | 0.63 | 0.62 |
| SVM (BOW) | 0.99 | 0.63 | 0.65 |
| SVM (TFIDF) | 0.98 | 0.64 | 0.64 |
| BERT | 0.79 | 0.75 | 0.74 |
| **RoBERTa** | **0.76** | **0.75** | **0.75** |

Table 1: Performance comparison of different models on ABSA tasks

sentiment counts, providing a yearly trend score. This approach emphasizes shifts toward positive or negative sentiment and allows for clearer trend visualization. However, it may also amplify the appearance of consistently declining sentiment if negative posts dominate over time.

Across the 12 areas, several trends emerge. First, annotated datasets tend to show more fluctuation in sentiment, with several areas experiencing noticeable changes over time, either becoming more negative (e.g., Uittamo, Runosmäki) or showing improvement (Soukka, Suikkila). In contrast, the predicted datasets generally show smoother and more stable sentiment trends, often leaning towards neutrality or slight negativity. This smoothing effect may be due to the model's tendency to generalize sentiment across larger sets of data, failing to capture more subtle changes over time.

As illustrated in Figure 4, the annotated sentiment for Harittu exhibits a general downward trend, beginning with a positive polarity and gradually declining over the years, eventually becoming predominantly negative. A similar trend is observed in the predicted dataset, though the sentiment remains more stable over time, hovering slightly below neutral. Overall, the trend indicates that public sentiment towards Harittu has remained negative over time.

Another interesting area according to the SA results is Kontula as presented in figure 5. The annotated sentiment for Kontula remains negative throughout, with only minor fluctuations, but stays relatively stable. Similarly, the predicted dataset reflects a consistent negative sentiment, showing little variation over time. Both datasets suggest that public sentiment towards Kontula has consistently remained negative with minimal changes. As another example, in the annotated data, sentiment for Varissuo(shown in Figure 6) fluctuates between slightly negative and neutral, with no strong or consistent trends emerging. The predicted sentiment remains stable, consistently neutral, with only minor fluctuations. Both datasets indicate that sentiment towards Varissuo has remained relatively unchanged over time, with no significant shifts. We recognize that the inherent tendency of users to share more negative experiences than positive ones on social media may indeed influence the sentiment trends observed in this study. Future analyses could benefit from additional controls or normalization methods to account for the naturally higher volume of negative feedback in social media contexts, which might give a more balanced view of sentiment trends over time.

## 4 Discussion

Naturally occurring online discussions offer an opportunity to study the sentiments of the general public about neighborhoods form a linguistic point of view. Understanding such sentiments is crucial for regional development, urban planning, and public policy. This study aimed to investigate sentiment toward 12 different neighborhoods in Finland by leveraging user-generated content from Suomi24.

Using a combination of classical machine learning models, such as SVM and Naive Bayes, alongside state-of-the-art models like BERT and RoBERTa, we applied ABSA to analyze the sentiment associated with each neighborhood. The RoBERTa model outperformed other methods, demonstrating its superior ability to predict sentiment accurately. We extended our research to analyze sentiment trends for a larger dataset of 32,183 data points, which offers insights into how public opinion evolves over time. Our results contribute to the growing body of knowledge in SA, particularly in the context of regional perceptions, and have practical implications for urban planning and policy making.

We conducted an error analysis using the best-performing model, RoBERTa, to investigate the misclassifications made on the test set. As previ-

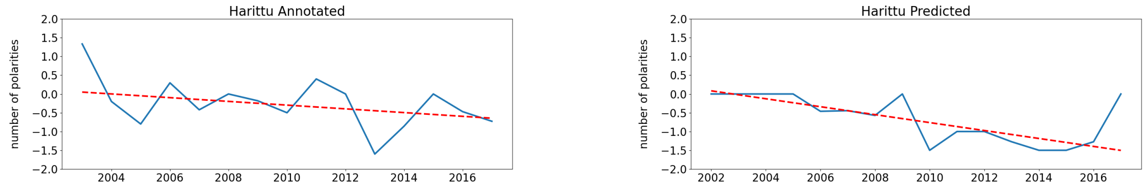



Figure 4: Harittu sentiment change overtime

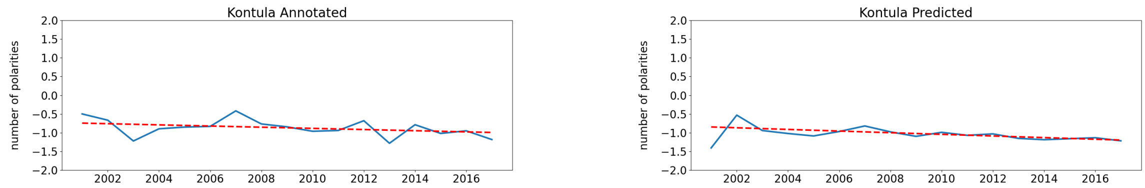



Figure 5: Kontula sentiment change overtime

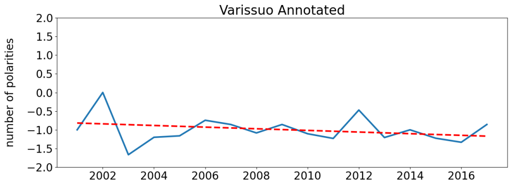
ously mentioned, the lemmatized version of the text was used for both model training and evaluation. However, in this section, we present examples in their original form to enhance readability and provide clarity on the context.

A significant number of misclassifications stemmed from ambiguous or multi-layered statements where the emotional tone was not immediately clear. The model tended to predict negative sentiment when it encountered critical or confrontational language, whereas human annotators perceived such content as neutral or factual. For instance, the sentence *"2 kysymystä mitä sä sitten riehut. vai uhkaako teidän perhettä <TAG>Matinkylä</TAG> syndrooma?"* is neutral while the model predicted a negative sentiment. Furthermore, the model struggled with indirect sentiments, such as rhetorical or sarcastic statements, which it frequently misinterpreted as negative when the intended sentiment was neutral or sarcastic. An example is the sentence *"Suomalaisen luonteen heikkous näkyy Eikö <TAG>kontula</TAG> ole oikeasti gårdsbacka ?"* which criticizes the "weakness" of the Finnish character but does not directly express negativity towards the area (Kontula). Lastly, some factual statements were misclassified, likely because the model identified subtle emotional cues that the annotators did not prioritize. For instance, *"Nykyään rakennetaan hometaloja ! Ei <TAG>kontulan</TAG> ostarilla ole vanhoja jykeviä kivitaloja"* can be interpreted as factual. The phrase "rakennetaan hometaloja" (meaning "nowadays they build moldy houses") might imply dissatisfaction with construction practices through sarcasm. Increasing the amount of training data could enhance the model's ability to handle these complexities more effectively.

The analysis of sentiment trends across the 12 neighborhoods reveals a range of patterns, with most neighborhoods showing either stable or slightly declining sentiment over time. Areas such as Leppävaara, Pihlajamäki, and Uittamo exhibit a subtle but consistent decline in sentiment, with annotated data indicating a shift from neutral or slightly positive to more negative sentiment. Matinkylä and Maunula remain largely neutral, with minimal fluctuations in sentiment across both annotated and predicted datasets, indicating a stable public perception. Myllypuro and Runosmäki show a predominantly negative sentiment, with Runosmäki demonstrating greater variability in the annotated data, while predicted data captures a more consistent negative trend. In contrast, Soukka and Suikkila display improvements in sentiment over time according to the annotated data, though the predicted data does not fully reflect this positive shift, suggesting a potential limitation in capturing more complex emotional changes.

## 5 Conclusion

This study presents a manually annotated dataset for ABSA tasks for studying different areas in Finland. As far as we know, there is no publicly available dataset for this specific domain in the Finnish language. The dataset consists of 3183 sentences taken from Suomi24. The dataset has been designed to perform various research tasks such as aspect extraction, and polarity detection. We believe this dataset would provide considerable significance in the research area of ABSA for social



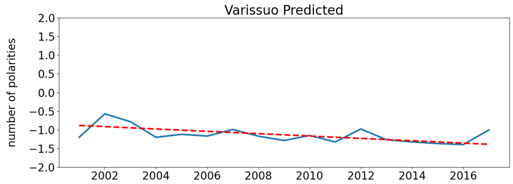




Figure 6: Varissuo sentiment change overtime

media text in the Finnish language. The dataset will be published publicly to be used by researchers in Finland. Moreover, our research addresses a significant gap by using SA to understand public perceptions of different areas in Finland. Utilizing data from Suomi24, one of the largest online forums in Finland, allows for the collection of unfiltered opinions, offering an authentic insight into how residents perceive their living environments. By applying state-of-the-art ABSA models like RoBERTa, our study enhances the analysis of regional sentiment. Additionally, tracking sentiment trends over time using 32,183 data points can provide valuable information for urban planners, policymakers, and social scientists. This combination of large-scale, time-sensitive data analysis and advanced machine learning techniques makes our research highly relevant for improving public policy and regional development. Future research could explore the relationship between sentiment and real estate pricing in Finland by analyzing how changes in public sentiment over time correlate with fluctuations in housing prices. By pairing SA with real estate data, it may be possible to predict price trends in specific areas or explain price shifts based on sentiment dynamics. A similar approach has been applied in other markets, such as the study by (Wang and Hui, 2017) which analyzed the predictive power of sentiment on market indicators like price and transaction volume in the housing market. Applying such a methodology to the Finnish housing market could provide new insights into how local sentiment influences real estate trends, potentially offering predictive power for pricing fluctuations across different areas.

Finally, a limitation of this study is the potential impact of dataset imbalance and prediction errors, which can affect the detection of complex sentiment trends in the larger predicted dataset. An imbalanced training set, with a prevalence of neutral sentiments, may lead the model to favor stable or neutral predictions. Thereby, smoothing out fluctuations that might otherwise reveal slight shifts in sentiment. Addressing these challenges in future work by employing techniques to balance sentiment classes or reduce prediction errors could improve the model's sensitivity to evolving sentiment dynamics across neighborhoods.

Moreover, we acknowledge that the number of posts may have varied significantly across the years, particularly as internet and social media usage expanded in Finland. However, while our focus was on examining sentiment trends rather than volume dynamics, future work could delve deeper into how this growth in user base and posting frequency influences sentiment patterns. Incorporating controls for yearly post volumes could offer a clearer view of whether observed sentiment shifts are affected by posting frequency or reflect genuine changes in public sentiment.

## Acknowledgments

This research was originally part of a project funded by the Turku Urban Research Program. Its primary goal was to investigate how neighbourhood sentiments expressed in public online discussions affect housing prices in Finland using digital and statistical data, including sources such as Suomi24 discussions.

We would like to extend our sincere gratitude to Professor Jozsef Mezei for providing invaluable support with technical challenges and suggestions in the writing process.

## References

Munir Ahmad, Shabib Aftab, Muhammad Salman Bashir, Noureen Hameed, Iftikhar Ali, and Zahid Nawaz. 2018. Svm optimization for sentiment analysis. *International Journal of Advanced Computer Science and Applications*, 9(4).

Twil Ali, Bidan Marc, Bencharef Omar, Kaloun Soulaimane, and Safaa Larbi. 2021. Exploring destination's negative e-reputation using aspect based sentiment analysis approach: case of marrakech destination on tripadvisor. *Tourism Management Perspectives*, 40:100892.

Ferdaous Benrouba and Rachid Boudour. 2023. Emotional sentiment analysis of social media content for mental health safety. *Social Network Analysis and Mining*, 13(1):17.

Kehao Chen and Guiyu Wei. 2023. Public sentiment analysis on urban regeneration: A massive data study based on sentiment knowledge enhanced pre-training and latent dirichlet allocation. *Plos one*, 18(4):e0285175.

Diogo Cortiz. 2022. Exploring transformers models for emotion recognition: A comparision of bert, distilbert, roberta, xlnet and electra. In *Proceedings of the 2022 3rd International Conference on Control, Robotics and Intelligent System*, pages 230–234.

Laleh Davoodi and József Mezei. 2022. A comparative study of machine learning models for sentiment analysis: Customer reviews of e-commerce platforms. pages 1–14.

Jacob Devlin. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Hai Ha Do, Penatiyana WC Prasad, Angelika Maag, and Abeer Alsadoon. 2019. Deep learning for aspect-based sentiment analysis: a comparative review. *Expert systems with applications*, 118:272–299.

Taqwa Hariguna and Vera Rachmawati. 2019. Community opinion sentiment analysis on social media using naive bayes algorithm methods. *International Journal of Informatics and Information Systems*, 2(1):33–38.

Rickard Hellström. 2022. Aspect based sentiment analysis (absa) for finnish language. Master's thesis, Aalto University.

Haosheng Huang and Georg Gartner. 2016. Using mobile crowdsourcing and geotagged social media data to study people's affective responses to environments. *European handbook of crowdsourced geographic information*, 385.

Zhengwei Huang, Jing Ouyang, Xiaohong Huang, Yanni Yang, and Ling Lin. 2021. Explaining donation behavior in medical crowdfunding in social media. *Sage Open*, 11(2):21582440211014520.

Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6280–6285, Hong Kong, China. Association for Computational Linguistics.

Akbar Karimi, Leonardo Rossi, and Andrea Prati. 2021. Adversarial training for aspect-based sentiment analysis with bert. In *2020 25th international conference on pattern recognition (ICPR)*, pages 8797–8803. IEEE.

Sayyida Tabinda Kokab, Sohail Asghar, and Shehneela Naz. 2022. Transformer-based deep learning models for the sentiment analysis of social media data. *Array*, 14:100157.

Krista Hannele Lagus, Minna Susanna Ruckenstein, Mika Pantzar, and Marjoriikka Jelena Ylisiurua. 2016. Suomi24: muodonantoa aineistolle.

Krister Lindén, Tommi Jauhiainen, and Sam Hardwick. 2023. Finnsentiment: a finnish social media corpus for sentiment polarity annotation. *Language Resources and Evaluation*, 57(2):581–609.

Yinhan Liu. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Pablo Martí, Leticia Serrano-Estrada, and Almudena Nolasco-Cirugeda. 2019. Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74:161–174.

Arun Meena and Tadinada Vankata Prabhakar. 2007. Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis. In *Advances in Information Retrieval: 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007. Proceedings 29*, pages 573–580. Springer.

Payal Mehra. 2023. Unexpected surprise: Emotion analysis and aspect based sentiment analysis (absa) of user generated comments to study behavioral intentions of tourists. *Tourism Management Perspectives*, 45:101063.

Nghia Luan Pham, Thang Viet Pham, et al. 2023. A data augmentation method for english-vietnamese neural machine translation. *IEEE Access*, 11:28034–28044.

Jin Rui. 2023. Exploring the association between the settlement environment and residents' positive sentiments in urban villages and formal settlements in shenzhen. *Sustainable Cities and Society*, 98:104851.

Marzieh Saeidi, Guillaume Bouchard, Maria Liakata, and Sebastian Riedel. 2016. Sentihood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods. *arXiv preprint arXiv:1610.03771*.

Dipti Sharma and Munish Sabharwal. 2019. Sentiment analysis for social media using svm classifier of machine learning. *Int J Innov Technol Exploring Eng (IJITEE)*, 8(9):39–47.

Pascal Soucy and Guy W. Mineau. 2005. Beyond tfidf weighting for text categorization in the vector space model. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1130–1135.

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307.

Amira Samy Talaat. 2023. Sentiment analysis classification system using hybrid bert models. *Journal of Big Data*, 10(1):110.

Feilong Tang, Luoyi Fu, Bin Yao, and Wenchao Xu. 2019. Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences*, 488:190–204.

Demircan Tas and Rohit Priyadarshi Sanatani. 2023. Geo-located aspect based sentiment analysis (absa) for crowdsourced evaluation of urban environments. *arXiv preprint arXiv:2312.12253*.

Anina Troya, Reshmi Gopalakrishna Pillai, Dr Cristian Rodriguez Rivero, Dr Zulkuf Genc, Dr Subhradeep Kayal, and Dogu Araci. 2021. Aspect-based sentiment analysis of social media data with pre-trained language models. In *Proceedings of the 2021 5th International Conference on Natural Language Processing and Information Retrieval*, pages 8–17.

Satuluri Vanaja and Meena Belwal. 2018. Aspect-level sentiment analysis on e-commerce data. In *2018 International conference on inventive research in computing applications (ICIRCA)*, pages 1275–1279. IEEE.

A Vaswani. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*.

Maria Vilkuna. 1989. Free word order in finnish.

Antti Virtanen, Jenna Kanerva, Rami Ilo, Jouni Luoma, Juhani Luotolahti, Tapio Salakoski, Filip Ginter, and Sampo Pyysalo. 2019. Multilingual is not enough: Bert for finnish. *arXiv preprint arXiv:1912.07076*.

Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. 2014. Dcu: Aspect-based polarity classification for semeval task 4. In *International Workshop on Semantic Evaluation (SemEval-2014)*, Dublin, Ireland.

Zixiao Wang. 2023. Emotion analysis-based decision support system for public perception evaluation in urban planning and design using social media text. *Soft Computing*.

Ziyou Wang and Eddie Chi-man Hui. 2017. Fundamentals and market sentiment in housing market. *Housing, Theory and Society*, 34(1):57–78.

Meylan Wongkar and Apriandy Angdresey. 2019. Sentiment analysis using naive bayes algorithm of the data crawler: Twitter. In *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pages 1–5. IEEE.

Qianwen Ariel Xu, Victor Chang, and Chrisina Jayne. 2022. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, 3:100073.

Dongyang Yan, Keping Li, Shuang Gu, and Liu Yang. 2020. Network-based bag-of-words model for text classification. *IEEE Access*, 8:82641–82652.