National College of Ireland

Sentiment Analysis of Ireland Housing Problem using Ensemble Learning and XAI Technique

MSc Research Project

Cloud Computing

Student Name: Basanti Pun

Student ID: 21130931

School of Computing

National College of Ireland

Supervisor: Dr. Shivani Jaswal

National College of Ireland
MSc Project Submission Sheet
School of Computing

Student Name:	Basanti Pun		
Student ID:	21130931		
Programme:	MSc in Cloud Computing	**Year:**	2023
Module:	Research Project		
Supervisor:	Dr. Shivani Jaswal		
Submission Due Date:	25/04/2023		
Project Title:	Sentiment Analysis of Ireland Housing Problem using Ensemble Learning and XAI Technique		
Word Count:	8546		
Page Count	22		

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Basanti Pun
Date:	25/04/2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Table of Contents

Sentiment Analysis of Ireland Housing Problem using Ensemble Learning and XAI Technique

Basanti Pun
21130931

Abstract

The problem of housing has become a critical issue in Ireland, and it is necessary to assess public sentiment towards it using sentiment analysis to locate areas for improvement. This research uses a combination of machine learning techniques, such as Logistic Regression, Support Vector Machine (SVM), and Decision Tree, to gain more accurate insights into the sentiments expressed in the data. The researchers also use an XAI technique called LIME to ensure transparency and interpretability. Amazon Sagemaker is used to perform the complete analysis for the research work. The data is collected from Twitter through different hashtags related to housing in Ireland. The primary aim of this study is to identify the location where most people face this problem. The ensemble technique utilizes classifiers, including logistic regression, SVM, and decision tree, for classification and then recognizes important or unknown factors' insights from the data with the help of XAI technique, which adds value to the sentiment analysis. The ensemble method achieves an overall classification accuracy of 70%. The objective of this research is to support future researchers in the fields of machine learning and sentiment analysis concerning housing issues.

1. Introduction

1.1 Background

The issue of housing affordability in Ireland has been a pressing concern for a prolonged period, and people are struggling to find affordable housing options. This problem has triggered a strong public response, with many individuals expressing their thoughts and sentiments through online platforms. Sentiment analysis can effectively analyse these public opinions and concerns. This approach involves utilizing various machine learning algorithms and natural language processing techniques to extract emotions and opinions from textual data. Sentiment analysis is gaining popularity in multiple domains, including politics and marketing.

In this research work, an approach is proposed for identifying emotions or sentiments of public through the textual form of data related to the housing problem of Ireland with the help of an ensemble of machine learning algorithms and XAI technique. The data used in this research project is collected from various different sources such as social media or news platforms or blogs etc. This research aims to recognize the area where the majority of people facing this problem based on the sentiment analysis and then we'll evaluate working of different ML algorithms in identifying the sentiments. The data present in textual format is pre-processed with NLP methods such as stemming, tokenization and part of speech tagging. An ensemble of different ML algorithms is used such as decision tree, SVM and LR to identify the sentiments. An ensemble learning is a method which includes grouping various algorithms together for increasing accuracy and efficiency of ML algorithms. Ensemble learning will be

ideal in this process of sentimental analysis as one ML algorithm will conceal the gaps of another.

Lalita et al. (2019) uses an ensemble approach for sentiment analysis on student comments in the faculty evaluation which achieves high accuracy in classifying the student sentiments. Although the study was able to achieve high accuracy, they fail to explain how they arrived at those predictions. Thus we need a transparent model which provides explanation for their decisions. We use XAI technique for expressing how these ML algorithms have worked in generating predictions. The XAI technique used in this particular research work is LIME. This XAI technique will help to maximize the interpretation ability and transparency of the model. It will also provide insight of data which will help in the process of sentimental analysis.

1.2 Motivation

The Ireland housing problem has gained significant attention in recent years, as the nation grapples with issues such as affordability, accessibility, and quality of housing. Policymakers, researchers, and the public alike are keen to understand and address the concerns associated with this problem. The analysis of sentiment on this topic can provide valuable insights into public opinion, facilitating the development of effective and informed policies and interventions.

The motivation behind using ensemble learning techniques in sentiment analysis lies in their potential to improve the accuracy and reliability of predictions by combining multiple models. By leveraging the strengths of diverse models and reducing individual weaknesses, ensemble learning can potentially produce more robust and accurate results compared to single-model approaches. In the context of the Ireland housing problem, this translates into a more reliable understanding of public sentiment, which can lead to better decision-making and resource allocation.

Explainable Artificial Intelligence (XAI) techniques are critical in enhancing the interpretability and trustworthiness of sentiment analysis models. As complex algorithms become more prevalent, ensuring that their decision-making processes are transparent and understandable is essential to maintaining public trust and fostering collaboration. Integrating XAI technique with ensemble learning models for sentiment analysis will not only provide insights into public opinion but also reveal the underlying reasons behind the predictions, making it easier to identify potential biases and address ethical concerns. Thus, the combination of ensemble learning and XAI technique aims to provide a more effective and responsible approach to understanding and addressing the Ireland housing problem.

1.3 Research Questions

1. How can ensemble learning techniques and explainable artificial intelligence (XAI) technique be applied to enhance the interpretability, trustworthiness, and effectiveness of sentiment analysis models for the Ireland housing problem?
2. How can sentiment analysis using ensemble learning and XAI technique help identify and address major challenges faced by various stakeholders in the Ireland housing problem and contribute to the formulation of effective policies and interventions?

1.4 Research Objectives

The main objectives of the report are outlined below -

1. To develop a sentiment analysis model utilizing ensemble learning techniques to accurately capture public opinion on the Ireland housing problem and to integrate explainable artificial intelligence (XAI) technique with the ensemble learning models to enhance interpretability, transparency, and trustworthiness of the sentiment analysis process.
2. To identify key features that contribute to the performance of ensemble learning models in sentiment analysis and to assess the impact of the findings from sentiment analysis using ensemble learning and XAI technique on policy formulation and intervention development for the Ireland housing problem.

This technical report follows the following structure. In section 2, we highlight and review the most relevant literature review for this project. Detailed explanations of the methodologies are provided in Section 3. Section 4 describes the design specification, including the Technologies & Frameworks, Architecture, and Proposed Model Functionality. In Section 5, we show how the proposed system was implemented. The evaluation is provided in section 6, which includes the results and discussion. The final section, Section 7, covers the conclusion of the research and future work.

2. Related Work

The problem of hunting for an affordable house has been a significant concern in Ireland. In the recent era, sentimental analysis is used for understanding the opinions and emotions of people on different problems such as housing. This section will provide a review of the related literature based on the concept of sentimental analysis and XAI techniques. In this time, the issue of having an adequate and adorable house has become a problem in many countries including Ireland. So, with the help of sentimental analysis, it will be helpful for the stakeholders and policy makers to understand the opinion of public and understand the area of concern.

Lee et al. (2015), presented that algorithm such as multinomial Naive Bayes for classification of text. When this model was compared with other algorithms such as SVM, NB or RF, this model outperformed with a F1 score of 0.811. this literature survey shows a gap that the author used very few algorithms for the classification purpose. The author should have explored some more algorithm in order to get improved results.

Nayak et. al (2016) gathered information of the year 2003 from the Yahoo finance and used this data into two different models in order to predict the trend of the stock market. One model used to make predictions based on daily activity of the stock market and another model makes monthly predictions. The dataset used for both the models was different. These datasets were modelled with many ML algorithms such as SVM, LR or decision tree. The author observed an accuracy of 70% when using the SVM algorithm. However, it not sure how vigorous this approach is in various situation or events which can impact the trend of stock market.

Tang et al. (2016) used memory network to tackle the aspect-based sentiment analysis task by storing the context of a given aspect in a memory network and measuring the degree of aspect word-context association using an attention mechanism, which has a faster computational rate than LSTM. All of these models, nevertheless, use static word vectors and are unable to dynamically change word meanings in response to context during model

training to account for words; many possible meanings. But this research work lacks in employing a sentence structure such as analysing results into a deep memory language.

Kaili et al. (2016) proposed a deep region and multi-label learning (DRML), a unified deep network. The DRML is a region layer that forces the learnt weights to capture the structural details of the face by forcing feed-forward functions to generate significant facial areas. The entire network is trainable from beginning to finish and automatically picks up representations that are resistant to local fluctuations. The main gap faced by this survey is that the author ensured that DRML can itself pick representation which is limited to the local instabilities, but is not able to explain the representation related with facial expression and features.

Chu et al. (2017) proposed a multi-level facial AU identification technique incorporating spatial and temporal characteristic. First, a CNN is utilized to extract the spatial representations, reducing the person-specific biases brought on by manually created descriptors (e.g., HoG and Gabor). No matter how long the input video sequences are, LSTMs are piled on top of these representations to simulate the temporal dependencies. A fusion network that further aggregates the outputs of CNNs and LSTMs produces a per-frame prediction of 12 AUs. Although, this literature survey also faces a gap as this research has probably not shown in standard hand-crafter attributes for example LBP, shape, HOG or gabor for the detection of AU.

Wu et al. (2017) introduced two models which are ELM (extreme learning machine) and RBF (radial basis function) with the help of SA approach for classification of aging defects, short circuit and partial shading in the PV system. The model proposed by them helps in increasing accuracy. Another model proposed by ensemble learning used some ML algorithms such as quadratic discriminate analysis, decision tree with entropy and this model identified short-circuit and partial shading with the accuracy of 97.67% (Kapucu 2021). Although, the proposed method may attain a good accuracy but the literature gap it faces is that it doesn't provide any explain ability for the action taken in the process of classification. By generating more insight into the process of decision making this model can enhances their trust and explain ability.

Hasani et al. [2017] presented a 3D Inception-ResNet design followed by an LSTM unit, which collectively extracts the spatial and temporal relationships within the facial pictures between various frames in a video sequence. This network also takes facial landmark points as inputs, stressing the relevance of facial components over facial regions, which might not be as important in producing facial expressions. The literature gap present in this study is that the proposed method is not compared with the existing model of this domain which can help in identifying any room of scope for the proposed model.

Meena et. al. (2018) showed a framework for extracting tweets and studying users' opinion on the SIoT. This framework permits to get feed of twitter. In the process of data cleaning and it's pre-processing it perceives the slang, remove stop words and applies lemmatization. After this process, the process of extensive sentimental analysis was based on fragment vector model (FVM), improved popularity classifier (IPC), sentiwordnet(SWNC) and a model combining the methods of SNWC and IPC. The result showed that the FVM model attains the best accuracy out of all which is 94.88%. however, this approach has to be

compared with other techniques. The literature gap faced by this study is that the author was restricted with a limited amount of tweet used for the analysis.

Alrumayyan et. al. (2018) in which they studied about the general behaviour of users and about the "food" venue of the Riyadh, Saudi Arabia. They suggested to people to share about their experiences more often from the food related places as related to other places. This study also limits up to the detection of general behaviour of customers of Riyadh, South Arabia. The research should fill this study gap by implementing their research in the other parts of the county also.

Bardhan et. Al (2019) used the tool of sentimental analysis for gender mainstreaming in the house management of slum area in Mumbai, India. In this study, the author used the concepts of sentimental analysis to classifying the gender-specific requirements of slum tenants and create a gender-sensitive policies. The result of this study showed that the concept of sentimental analysis can also be used in the gender-specific problems in the housing management. The main limitation faced by this analysis is that in the cooperative society it is mandatory to have at least two females.

Nursal et. al. (2019) introduced an integrated multi-feature decision making and sentimental analysis methods for enhancing the project based on the residential housing purchase. In this study also, concepts of sentimental analysis are used for classifying opinions of public on the project related to house purchasing and combined it with multi-feature decision making to select the best possible project. The results provides that this concept may enhance the accuracy of the process of decision making and selecting a residential housing project. The gap present in this literature is that the study is limited for the method of integrated multi-feature decision making and sentimental analysis related for the issue of purchasing house.

Al-Agha et al. (2019) conducted a multi-level analysis of political sentiments using Twitter data in the context of the Palestinian-Israeli conflict. The study used sentiment analysis to understand the political sentiments of Twitter users and identified the areas of most concern. The results showed that sentiment analysis can be an effective tool for understanding the public opinion on political issues. The gap faced by this research work is that it lacks the sample representative of data obtained from twitter for the task of sentimental analysis.

Shaeeali et. Al. (2020) uses AI algorithms and technique for sentimental analysis on FDS in an analysis on social media customer review analytics on FDS. Lexicon, support vector machine (SVM), natural language processing (NLP), and text mining were the four AI algorithms analysed and compared. NLP came in second with 71.67 percent accuracy, followed by SVM with 69.70 percent accuracy, and text mining with 67.94 percent accuracy. As a result, the lexicon-based method outperforms machine learning (SVM) algorithms. The literature present in this study is that the author could conduct an in-depth analysis on the limitation of sentimental analysis on the review of social media customers by utilizing the method based on lexicon.

Li et. al (2020) uses SHAP technique for interpretation of attribute selection in their system of visual analysis in the clinical data. They presented local descriptions of grouped and individual instances. This technique permits comparison of different model by presenting

similarities of different machine learning algorithms based on feature extraction. The technique proposed by the author has a major limitation that the main objective of this research was on describing the use of SHAP value in the interpretability of the model and the outcome of the result was not able to present a quantitative estimation of SHAP value method.

Bhattarai et. al. (2020) utilized the concepts of TM's conjunction clauses for representing and learning the known classes and also in establishing a mechanism based on novelty scoring. Further, in next year the author adopted same technique for capturing binary semantic feature and vocabulary for predicting the authenticity of news and evaluating the reliability of the fake news. The gap present in this survey is that the proposed method is not compared with the performance of existing model for detection of false news as it would help to recognize the future scope and pros and cons of the proposed method.

Shamshirband et. al. (2020) analysed the utilization of deep belief networks (DBN), AE (auto-encoders), CNN and RNN in the domain of healthcare. They highlighted different problems and challenges with the concept of deep learning model in healthcare and into the specific insights into the applicability and accuracy of DL models. The gap present in this analysis is that it the study lacks its attention on the ethical suggestion of utilizing DL models in the domain of healthcare.

Chromik et al. (2021) proposed different design principles for XUI, which include- i. making a group of images and texts to create a better understanding and communicate with efficiency; ii. Provides iterative or hierarchical or hierarchical functionality which effectively permits prior description and follow-ups; iii. Several methods of explanation and modalities help to triangulate insights; and iv. Regulate descriptions to the user's intellectual model and context. The above-mentioned principles can improve the application of XAI from traditional methods output to an easy model. However, the proposed method has some research gap as the author missed the related applied research from adjacent communities of XAI outcome and inside of the ACM.

Sarkodie et al. (2021) proposed several regression models with the goal to provide assumptions of COVID-19's growth and mortality, considering environmental, health, and socio-economic data like air temperature, diabetes prevalence index, age, GDP per capita, etc. Despite the fact that the models take a variety of data points into account, including many different cities, the selected time period (1 January to 11 June 2020) may not be long enough to provide an accurate prediction. In the work of, a neural network architecture is presented which is based on a trained Long Short-Term Memory (LSTM) network. This network is designed to predict the upcoming daily COVID-19 cases. However, this study lacks and provides limited resources in terms of data availability, coverage and unequal distribution of data.

El-kassa et.al. (2021) discover that rationalization can be split into two different groups that are abstractive and extractive. In abstractive, rationalization is a generative work which creates the new sentences or rewrites the existing words. But in extractive rationalization, only the important sentences or features from the existing data are taken out as rationale for supporting the process of prediction. The main challenge faced by the research is most of researcher's main objective is on the extractive approach, it is important to pay more

attention on the research efforts to introduce and enhance the system of summarization which is based on the hybrid and abstractive approaches.

Balcilar et. al. (2021) in which they analysed the role of sentimental analysis on housing marketing prediction in US. The author used the concept of sentimental analysis to evaluate the monetary sentimental of public and its effect in the housing marketing. The author concludes that economic sentimental is a crucial factor for highlighting the gender-specified problem in housing management. The literature gap present in this approach is that the author tested this approach on a particular dataset which is related to housing market of US. But the author should have used more diverse and large datasets from various regions in order to get improved results.

Ahmed et.al. (2021) conducted a study on the detection of hate comments in Bengali language. The author collected the data from the comment section of Facebook which was related to hate. A total of 1300 comments in Bengali. The main aim of this research work is to use the method of binary classification in order to detect hate language from the comment on Facebook. When the author trained the model with ML algorithms like SVM and NB and for the feature selection used the technique of TF-IDF then get an accuracy of 72%. The literature gap faced by the approach is that the performance of this model doesn't satisfy much as the performance was conducted on a small dataset and it is not sure not this method will work on other datasets in the field of NLP as it is only being evaluated on dataset of Bengali language.

Minovski et. al. (2021) proposed a non-network intrusive model based on ML algorithms for anticipate the through in a dependent 5g network which is based on network slice. For understanding the process of decision making from the root causes analysis, a model based on the ML algorithm of decision tree is used and results in the accuracy of 93% while using an ANN for the optimal method used for the processing of data in tabular format. The gap present in this study is that it is only restricted to depended 5g which is based on network slice and is not employed on other 5g networks.

Chazette et. al. (2022) proposed a study of trade off as a subpart of requirement analysis. this study discusses a trade-off between the explanibility factor and other factors like user friendly, data load or user interface design are a part of requirement analysis. The author stated that, while setting up all the requirements, the importance of explaining the working and behaviour of an ideal system is as important as explaining models' consequences and uncertainty. With the help of this approach, consequences caused due to the unusual behaviour of the system can be reduced. The gap faced by this review is that the author only provides a theoretical argument of relation in requirement analysis which is not sufficient to authenticate models' efficiency practically.

Alicioglu et. al. (2022) conducted a survey based on visual analytic. The main topic of this survey was VA for the XAI techniques which are better for interpreting neural networks. In this survey author coved the current situation, limitations and future scope. The limitation faced by this research is that model does not provide any explanation which reduce the transparency or trust of the model.

Vasconcelos et. Al (2022) highlights the gaps in cognitive engagement will continue if the methods of XAI remain difficult to understand, as people are more likely to choose to

involve with descriptions or easily acceptable to AI by weighing the cognitive costs. However, there is gap in this research that the paper proposed is not able to address the effects and causes of AI explanation on the trust also the modulating and measuring of trust is considered to be out of the scope of the research.

Table 1

Paper Title	Year	Objective and Methodology	Research Gap
Mining social media streams to improve public health allergy surveillance.	2015	The study presented that algorithm such as multinomial naive Bayes for classification of text. When this model was compared with other algorithms such as SVM, NB or RF, this model outperformed with a F1 score of 0.811.	this literature survey shows a gap that the author used very few algorithms for the classification purpose. The author should have explored some more algorithm in order to get improved results.
Aspect level sentiment classification with deep memory network.	2016	The author employed a memory network with attention mechanism for aspect-based sentiment analysis, which is computationally faster than LSTM.	This research work lacks in employing a sentence structure such as analysing results into a deep memory language
Facial expression recognition using enhanced deep 3D convolutional neural networks.	2017	The author presented a 3D Inception-ResNet design followed by an LSTM unit, which collectively extracts the spatial and temporal relationships within the facial pictures between various frames in a video sequence.	The literature gap present in this study is that the proposed method is not compared with the existing model of this domain which can help in identifying any room of scope for the proposed model.
Learning spatial and temporal cues for multi-label facial action unit detection	2017	The author presented a multi-level facial AU identification technique that uses spatial and temporal characteristics. LSTMs are added to these representations to model temporal dependencies, and a fusion network produces per-frame predictions of 12 AUs regardless of the input video length.	This research has probably not shown in standard hand-crafter attributes for example LBP, shape, HOG or gabor for the detection of AU.
An intelligent fault diagnosis approach for PV array based on SA-RBF kernel extreme learning machine	2017	The author introduced two models which are ELM and RBF with the help of SA approach for classification of aging defects, short circuit and partial shading in the PV system. The model proposed by them helps in increasing accuracy	it doesn't provide any explainability for the action taken in the process of classification. By generating more insight into the process of decision making this model can enhances their trust and explainability.
Evaluating twitter data to discover user's perception about social internet of things	2018	The framework includes data cleaning, pre-processing and the process of extensive sentimental analysis based on fragment vector model (FVM), improved popularity classifier (IPC), sentiwordnet(SWNC) and a model combining the methods of SNWC and IPC.	The literature gap faced by this study is that the author was restricted with a limited amount of tweet used for the analysis.

Improving residential housing project purchase by using integrated multi-attribute decision making and sentiment analysis technique.	2019	The author proposed a multi-feature decision-making and sentiment analysis method to improve residential housing purchase projects, which yielded more accurate results.	The gap present in this literature is that the study is limited for the method of integrated multi-feature decision making and sentimental analysis related for the issue of purchasing house.
Human-XAI interaction: A review and design principles for explanation user interfaces	2021	Various principles of XAI methods for improving application of XAI on the existing model.	The author did not consider the applied research with the adjacent group of the result of XAI.
Bangla Online Comments Dataset	2021	The main aim of this research work is to use the method of binary classification in order to detect hate language from the comment on Facebook. the model is trained with ML algorithms like SVM and NB and for the feature selection used the technique of TF-IDF.	The performance of this model doesn't satisfy much as the performance was conducted on a small dataset and this method might not work on other datasets in the field of NLP as it is only being evaluated on dataset of Bengali language.
Throughput prediction using machine learning in lte and 5g networks	2021	The author proposed a non-network intrusive model based on ML algorithms. A model based on the ML algorithm of decision tree is used and results in the accuracy of 93% while using an ANN for the optimal method used for the processing of data in tabular format.	The gap present in this study is that it is only restricted to depended 5g which is based on network slice and is not employed on other 5g networks.
Measuring the Novelty of Natural Language Text using the Conjunctive Clauses of a Tsetlin Machine Text Classifier	2021	The article employs TM's conjunction clauses to represent and learn known classes, establish a novelty scoring mechanism, and capture binary semantic features and vocabulary to predict news authenticity and evaluate fake news reliability.	The gap present in this survey is that the proposed method is not compared with the performance of existing model for detection of false news as it would help to recognize the future scope and pros and cons of the proposed method.
A review on deep learning approaches in healthcare systems : Taxonomies, challenges, and open issues	2021	Deep belief networks (DBN), AE (auto-encoders), CNN and RNN are utilized in the domain of healthcare. Author gives insights into the applicability and accuracy of DL models	The gap present in this analysis is that the study lacks its attention on the ethical suggestion of utilizing DL models in the domain of healthcare.
Global effect of city-to-city air pollution, health conditions, climatic & socio-economic factors on COVID-19 pandemic	2021	A neural network architecture is presented which is based on a trained Long Short-Term Memory (LSTM) network. This network is designed to predict the upcoming daily COVID-19 cases	This study lacks and provides limited resources in terms of data availability, coverage and unequal distribution of data.

In this research work, the proposed method of various researchers has been discussed with their literature gaps. The research was primarily based on the sentimental analysis for the housing problem but further the use of sentimental analysis in different domain is discussed in the above-mentioned work. Various researchers presented their work in this domain and use different machine and deep learning model for the evaluation of their work. Also, the literature gap for each work is also being discussed for the improvement in the future work. Furthermore, this analysis will be beneficial for the students and future researchers to the related domain.

3 Research Methodology

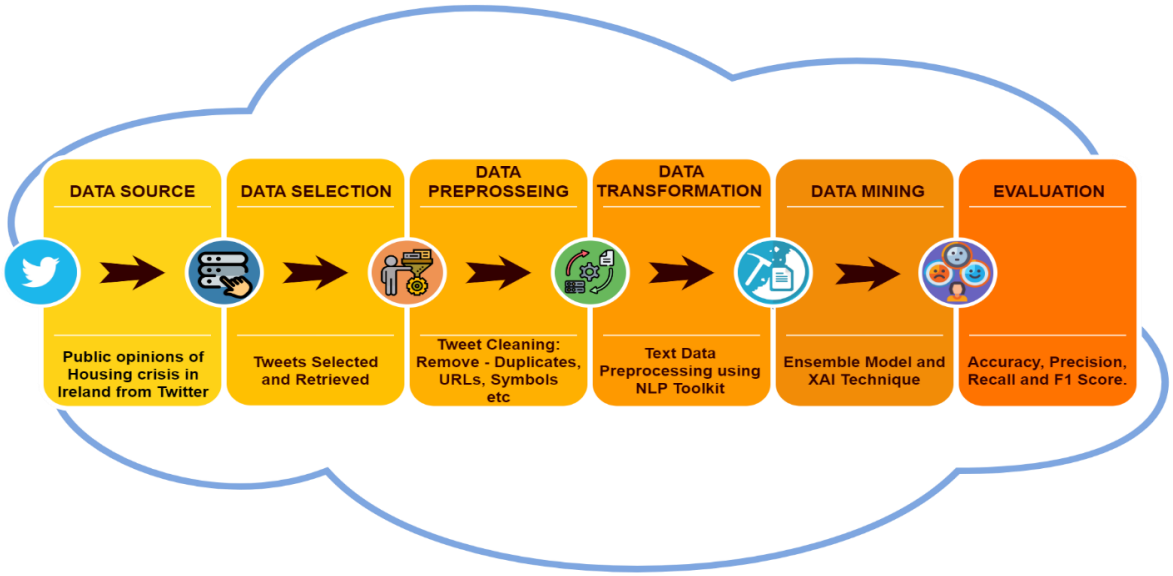

Figure 1. KDD process flow for Sentiment Analysis on Ireland's housing problem

This following section demonstrates the model used in this analysis to investigate the sentiment of people of Ireland regarding the issue of housing by using method of ensemble learning and XAI. The main aim of this project work is to generate a model based om sentiment analysis which is interpretable and precise for the stakeholders. The methodology is split up into sub-sections:

1. Data extraction and pre-processing
2. Data Transformation and feature extraction
3. Creating ensemble learning model
4. Implementation of XAI technique
5. Evaluation and validation of model

3.1 Data extraction and Pre-processing

The primary step in this research work includes extracting related data from different sources. In this analysis, data is collected from Twitter using hashtags related to housing problems in Ireland. The data will be gathered with use of web scraping and APIs which is provided by twitter. After the process of data collection, the raw data will go through the process of data

pre-processing to check the quality of data and relevance. The following steps are included in data pre-processing:

1. Eliminating identical entries.
2. Eliminating unrelated content such as non-English text or advertisements etc.
3. Tokenization: dividing text into separate words
4. Eliminating stop words: removing some common wors such as ‘is’, ‘and’, ‘the’ and many more.
5. Steaming and lemmatization: dropping words to their root or base forms in order to improved representation of text (for instance: ‘running’ to ‘run’)
6. Eliminating URLs, numbers, special characters: cleaning the text to hold characters present in alphabetical manner.
7. Translating text into lowercase: this step ensures dataset’s uniformity.

After the process of data cleaning and pre-processing, the pre-processed data is kept in structured way such as JSON, CSV for further study.

3.2 Feature Extraction

In this step, the text data is converted into such a format which can be easily used by ML classifiers, different methods of feature extraction will be implemented which includes:

1. BoW (Bag of Words): a representation which is use to count frequency of words inside a text.
2. TF-IDF (term frequency-inverse document frequency): an illustration which evaluates a word’s importance inside a text which is based on their frequency and shortage across the documents.
3. Word embeddings (GloVE, Word2Vec): a vector depiction which captures semantic relationships between the text.

The following features will be pull out from the cleaned and pre-processed data and is used as input for the model based on ensemble learning.

3.3 Creating ensemble learning model

The model of sentimental analysis will be generated by using methods of ensemble learning to improve its performance. The method of ensemble learning associations the assumption of various base models to generate robust and accurate output. In this analysis, multiple base model which includes random forest, naïve bayes, logistic regression SVM and DL models such as recurrent neural network and convolutional neural networks will be trained on the data with extracted features.

The methods of ensemble learning used in this analysis contains:

1. Bagging: a method which is used to train various base models on multiple subsets of the training data and merge the assumption by averaging or voting.
2. Boosting: a method which is used to train a sequence of base model, repetitively changing training data weight which is based on the performance of prior models.
3. Stacking: a method which associates the prediction of various base model by using a high-level or meta-level model.

The ideal combination of ensemble methods and base models will be resolute by cross-validation and experimentation.

3.4 Implementation of XAI technique

XAI methods is employed in order to improve interpretability of model based on sentimental analysis. The XAI method offers insight into the process of decision-making permitting stakeholders to learn and trust its assumption. The different XAI methods used in this analysis is:

1. Important feature analysis: recognizing most significant features in prediction of model.
2. Partial dependence plots: envisaging the association between output of model and a feature.
3. LIME (local interpretable model-agnostic explanations): describing every assumption by resembling the model with a local interpretable model.
4. Counterfactual descriptions: creating alternative input example which would have increase various predictions, helping stakeholders and learn the condition under which a model change its decision.

The implementation of XAI methods will permit stakeholders to gain an in-depth understanding of nature of model and the factors driving the perdition of sentiments.

3.5 Evaluation and Validation of Model

To calculate performance of model based on sentimental analysis, different evaluation metrics is used such as:

1. Accuracy: the ratio of correct assumption to the total number of samples.
2. Precision: the ratio of true positive assumptions to the total positive assumption.
3. Recall: the ratio of true positive assumption to the real positive samples.
4. F1-Score: the harmonic mean value of recall and precision, which provides a balance between both the metrics.

The performance of the model will be calculated by the use of k-fold cross validation, in which the dataset is split into k-subset, the model is trained and tested k number of times, every time use a different subset as test set. This method helps to certify that performance of model is independent of a training-test split and is consistent.

4 Design Specification

The presented framework for the sentimental analysis of the housing problem in Ireland with the help of ensemble learning and XAI technique will include following component:

4.1 Techniques and Frameworks

The model influences ensemble learning and XAI (explainable artificial intelligence) technique to conduct the process of sentimental analysis on the housing problem in Ireland. The method of ensemble learning helps in enhancing predictive accuracy and provides stability by grouping result of various base models. The techniques of XAI are used to learn and understand the process of decision-making of model and validates trust and transparency of the model.

The framework used in the project is:

1. Scikit-learn: it is ML library based on python which offers a range of unsupervised and supervised learning methods which is consist of methods such as bagging, random forest and gradient boosting.
2. LIME: LIME or local interpretable model agnostic explanation is a technique which is used to explain the prediction made by any of the classifier by resembling it with a locally interpretable model.

4.2 Architecture

The implementation of the model is generated on Amazon SageMaker, by using its ability to scale and manage ML based models. For the development and execution of the model, a jupyter notebook is built inside the Amazon SageMaker. The model's architecture includes following components:

1. Data collection: assembling textual form of data associated to the housing problem of Ireland from twitter
2. Data pre-processing: the assembled data gets cleaned and pre-processed in order to eliminate unrelated data, inconsistencies and noise.
3. Feature extraction: converting textual form of data into numerical form, such as vector of TF-IDF can be put up into the ML based models.
4. Model Training: training of various base models by using methods of ensemble learning, which includes boosting, stacking and bagging.
5. Model evaluation: evaluating the performance of model based on ensemble learning with the help of evaluation metrics such as precision, accuracy, F1scoreand recall.
6. Model interpretation: employing methods of XAI, for example LIME, to learn and explain the assumptions generated by the ensemble model.

4.3 Proposed Model Functionality

The present model based on ensemble learning associates' strengths of different base models in order to generate a stable and more accurate sentimental analysis model. Following ae some function of the model:

1. Every base model is trained on randomly different subset of the training data by the use of methods such as boosting and bagging.
2. The prediction generated by base model is combined with the help of methods like averaging or majority voting in order to generate the final result of ensemble model.
3. The ensemble model is then calculated on a testing dataset to regulate model's generalizability and performance.
4. The methods of XAI for example LIME, is employed to the prediction generated by ensemble model which offers interpretable reasons to the process of decision-making. This will make sure that the prediction generated by the model is trustworthy and transparent.

In conclusion, the design specification of this project includes the methods of ensemble learning and XAI technique in order to generate an interpretable and robust model of sentimental analysis based on the housing problem of Ireland. The implementation id creates on the Amazon SageMaker platform and influences popular framework such as Scikit-learn and LIME for the development and interpretation of the model.

5 Implementation

The implementation of proposed system is done in following ways:

1. Transformation of word
2. Removing duplicate entries
3. Eliminating unrelated content
4. Tokenization and stop-word removal
5. Lemmatization and stemming
6. Extraction of features

The output produced after implementation of model includes

Data transformation:
Raw data is collected from different social media sources, the extracted data is cleaned and pre-processed ensuing that the dataset includes related information and then converted into numerical representation as it is used as input in ML classifiers.

Code:
The code has been generated on python in jupyter notebook on the amazon SageMaker. The python code includes all the steps from extraction of data to data pre-processing, training, testing, evaluation and interpretation of using XAI and ensemble methods.

Model generated:
Various base models are trained with the use of ensemble methods such as boosting, bagging, and stacking. The main ensemble model was generated by collecting base models' prediction to enhances stability and accuracy of model.

Evaluation of result and prediction explainable:
The proposed models were evaluated on the basis of few evaluations and metrics to determine efficiency of model and with the help of XAI method such as LIME, explains the prediction result generated by model and also ensures trust and transparency of model.

The implementation of proposed system used following languages and tools:

Python:
The main programming language used for generating code. It is mostly selected because of its wide choices of libraries and user friendly in the task related to data analysis and ML.

Amazon SageMaker:
This platform is based on cloud used for creating, training and deploying ML algorithm at scale. Within the SageMaker, the jupyter notebook was generate which is used to support the progress and execution of code.

Scikit-learn:
It is a library of python which is used for the employment of ensemble learning methods and other task related to pre-processing and evaluation.

LIME:
It is a library in python which is used to generate interpretable explanations for the prediction generated by ensemble model.

6 Evaluation

6.1 Result of Data analysis

To determine the location having maximum housing problem in Ireland, the pre-processed dataset was observed. In a dataset of housing tweets, all unigram frequencies have been evaluated. Each type of sentiment work cloud either neutral, negative or positive is evaluated and showed below. For a better understanding of the nature of individual form of sentiment and how relays on the housing problem, few of data visualization tool is used.

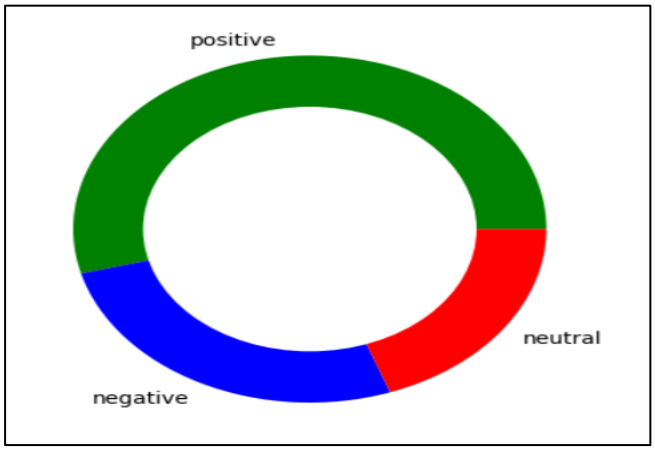

Figure 1 pie chart

The figure 2. is a pie chart which displays the proportion of each sentiment of tweet. The plot demonstrates the positive tweets has the maximum proportion and on the other hand, both negative and neutral tweets show equal amount of proportion.

A word cloud that is a graphical representation that displays the frequency of each word used in a collection of text documents. The size of each word is proportional to its occurrence in the dataset. Sentiment analysis is a technique used to determine the emotions or attitudes expressed in the text, whether positive, negative, or neutral. An all-encompassing word cloud can provide valuable insights into the most frequently used words that contribute to the expressed sentiment.

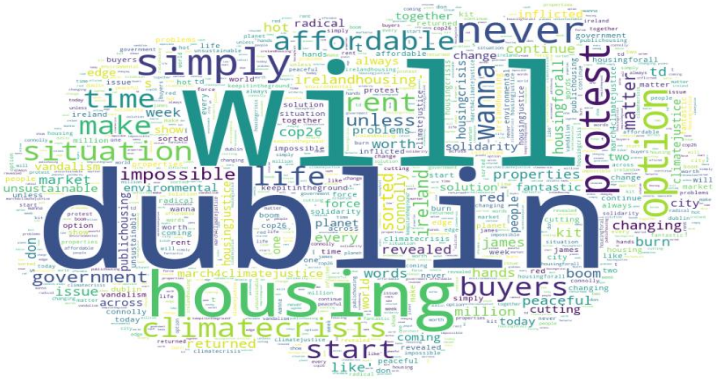

Figure 2 WordCloud for all tweets

In figure 3, the WordCloud of housing related tweets revels words such as “Dublin”, “housing”, “affordable”, “protest” and “will” which are frequently used in the tweets related to housing.

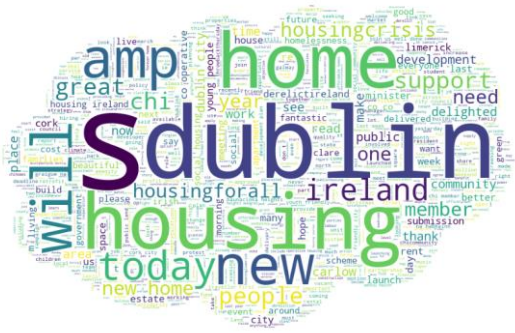

Figure 3 WordCloud for positive sentiment

In figure 4, the WordCloud for the positive type of sentiment discloses that “dublin”, “housing”, “home”, “housing for all”, “people” and “development” are frequently used in the tweets with the positive sentiment.

Figure 4 WordCloud for negative sentiment

In figure 5, the WordCloud of negative sentiment displayed words such as “housing”, “homeless”, “housingcrisis”, “crisis”, “Ireland” and “government” are frequently used in the tweets having negative sentiment.

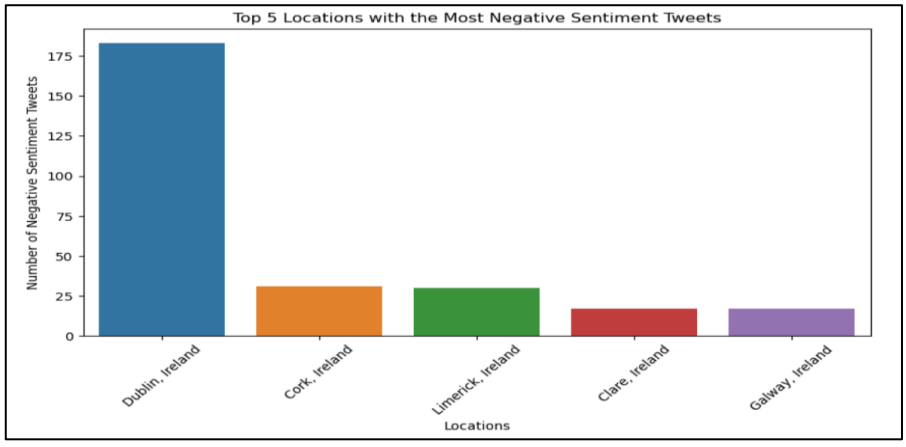

Figure 6 Top 5 locations with most negative sentiment tweets

Dublin topped the chart with the highest negative sentiment tweets followed by Cork, Limerick, Clare and Galway respectively.

6.2 Model's performance analysis

The experimental method for identifying housing problem of Ireland by the use of ensemble by using techniques such as bag of words, word embedding and TD-IDF on ensemble classifier. Different metrics of evaluation is used to compare the efficiency of ensemble classifier which is consist of ML model such as SVM, logistic regression and decision tree a then the proposed model is trained.

Table 2 Classification Report for various ML models

Model	accuracy	precision	recall	f1-score
Logistic Regression	0.64	0.68	0.64	0.59
Decission Tree	0.61	0.62	0.61	0.61
Ensemble Model(SVM,Logistic Regression and Decision Tree)	0.67	0.66	0.67	0.66
Ensemble Model(SVM,Logistic Regression and Decision Tree) with hyperparameter tuning	0.70	0.69	0.70	0.69

After the process of training, the model has been evaluated on testing set and generate a classification report. even though accuracy has a huge variety of application, it's not always that the ideal performance measure to use, especially in the case when the class of target variable in a dataset is unbalanced. So, few other performance metrics is used like F1 score which shows efficiency of model by using recall and accuracy. In Figure 7, different models were evaluated and their performances were compared through classification reports. The results indicate that the ensemble model, which underwent hyperparameter tuning, outperformed the other models with the highest accuracy score of 0.70. Additionally, it achieved the highest precision, recall, and f1-score of 0.69, 0.70, and 0.69, respectively.

6.3 Model interpretation by LIME

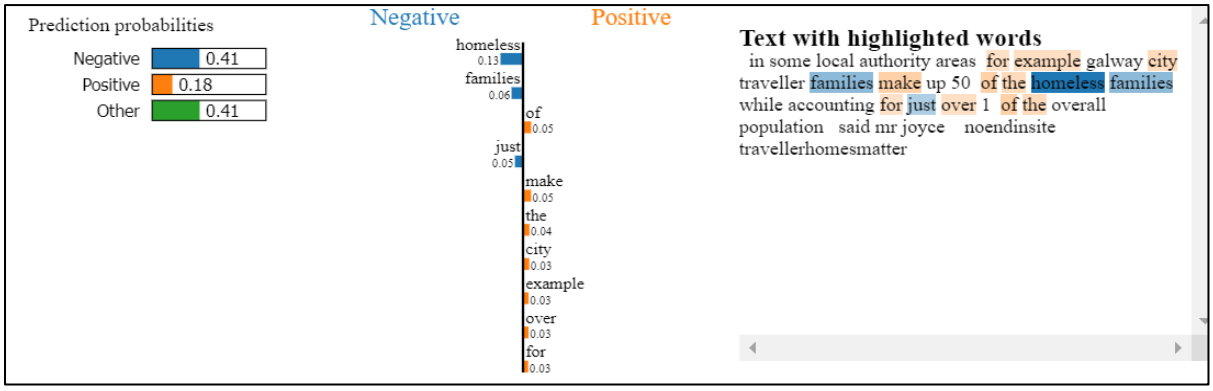

Figure 7 LIME interpretation

The figure 8 demonstrates that negative sentiment has 0.41 negative sentiment and 0.18 positive sentiment and 0.41 as other prediction probabilities. For the positive sentiments

“make”, “example”, ”over” shows and “city” contribute to the prediction probability. In the case of negative sentiment “homeless” has maximum range of 0.13 and words such as “family” and “just” have range of 0.06 and 0.05 respectively.

6.4 Discussion

The main goal of the project was to create models for sentiment analysis that were both accurate and efficient. After conducting a thorough literature review and identifying current limitations and gaps in the field. It was found that ensemble learning, which involves combining multiple models, and the XAI technique for interpretability were areas that had been less explored. The research work utilized ensemble learning with hyperparameter tuning. To fine-tune the hyperparameters, I used the GridSearchCV() function to create a grid of parameters, which was then used to fit the VotingClassifier to the training data. The model's performance was evaluated through cross-validation, and the best_params_ attribute of the GridSearchCV object was used to extract the optimal hyperparameters. I compared the performances of Logistic Regression, Decision Tree, Ensemble model, and Ensemble model with hyperparameter tuning, and found that the best results were obtained with the Ensemble model with hyperparameter tuning, which achieved an overall accuracy of 70%. The precision, recall, and f1-score scores were 69%, 70%, and 69%, respectively. To achieve transparency in regards to the sentiments of the tweet, LIME was used to generate a prediction probability for the tweets based on the test. This allowed to interpret the tweets and understand the sentiments expressed in them. The project also included identifying places with the highest negative sentiment related to Ireland's housing problem. By using Ensemble learning and the XAI technique Dublin was identified as the place with the highest number of tweets with negative sentiments, followed by Cork, Limerick, Clare, and Galway.

7 Conclusion and Future Work

The analysis carried out highlights the importance of using ensemble along with XAI technique to improve interpretability in machine learning and enhance the accuracy of sentimental analysis. The study also reveals the value of practitioner implicates in gaining insights into public opinion about the housing problem, which can be beneficial for urban planners, housing stakeholders, and policymakers to make informed decisions and develop effective solutions to address the housing crisis.. The evaluation of different models showed that the ensemble model with hyperparameter tuning performed better than other models, achieving an accuracy of 70%. LIME, an XAI method, was used to explain the predictions generated by the ensemble model, ensuring transparency and trust in the decision-making process. The results of this study will be useful for future researchers who can explore various ensemble methods and XAI techniques to improve accuracy. It would be intriguing to observe the performance of the models on a significantly larger labelled dataset as it may lead to superior outcomes. Additionally, it would be interesting to explore the implementation of deep learning techniques within the scope of this project as the results derived from this approach would also be insightful. Additionally, the study highlights the importance of combining different data sources and features to improve the performance of sentimental analysis.

References

Ahmed, M.F., Mahmud, Z., Biash, Z.T., Ryen, A.A.N., Hossain, A. and Ashraf, F.B., 2021. Bangla Online Comments Dataset. *Mendeley Data*, *1*.

Al-Agha, I. and Abu-Dahrooj, O., 2019. Multi-level analysis of political sentiments using Twitter data: A case study of the Palestinian-Israeli conflict. *Jordanian Journal of Computers and Information Technology*, *5*(3).

Alicioglu, G. and Sun, B., 2022. A survey of visual analytics for Explainable Artificial Intelligence methods. *Computers & Graphics*, *102*, pp.502-520.

Alrumayyan, N., Bawazeer, S., AlJurayyad, R. and Al-Razgan, M., 2018, March. Analyzing user behaviors: a study of tips in Foursquare. In *5th International Symposium on Data Mining Applications* (pp. 153-168). Cham: Springer International Publishing.

Balcilar, M., Bouri, E., Gupta, R. and Kyei, C.K., 2021. High-frequency predictability of housing market movements of the United States: the role of economic sentiment. *Journal of Behavioral Finance*, *22*(4), pp.490-498.

Bardhan, R., Sunikka-Blank, M. and Haque, A.N., 2019. Sentiment analysis as tool for gender mainstreaming in slum rehabilitation housing management in Mumbai, India. *Habitat International*, *92*, p.102040.

Bhattarai, B.; Granmo, O. and Jiao, L. (2021). Measuring the Novelty of Natural Language Text using the Conjunctive Clauses of a Tsetlin Machine Text Classifier. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,* ISBN 978-989-758-484-8; ISSN 2184-433X, SciTePress, pages 410-417. DOI: 10.5220/0010382204100417

Chazette, L., Brunotte, W. and Speith, T., 2021, September. Exploring explainability: a definition, a model, and a knowledge catalogue. In *2021 IEEE 29th international requirements engineering conference (RE)* (pp. 197-208). IEEE.

Chromik, Michael, and Andreas Butz. "Human-xai interaction: A review and design principles for explanation user interfaces." In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part II 18*, pp. 619-640. Springer International Publishing, 2021.

Chu, W.S., De la Torre, F. and Cohn, J.F., 2017, May. Learning spatial and temporal cues for multi-label facial action unit detection. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 25-32). IEEE.

El-Kassas, W.S., Salama, C.R., Rafea, A.A. and Mohamed, H.K., 2021. Automatic text summarization: A comprehensive survey. *Expert systems with applications*, *165*, p.113679.

Hasani, B. and Mahoor, M.H., 2017. Facial expression recognition using enhanced deep 3D convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 30-40).

Kaili, Z., Chu, W.S. and Zhang, H., 2016. Deep region and multi-label learning for facial action unit detection. In *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3391-3399).

Lee, K., Agrawal, A. and Choudhary, A., 2015, August. Mining social media streams to improve public health allergy surveillance. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (pp. 815-822).

Li, R., Shinde, A., Liu, A., Glaser, S., Lyou, Y., Yuh, B., Wong, J. and Amini, A., 2020. Machine learning–based interpretation and visualization of nonlinear interactions in prostate cancer survival. *JCO Clinical Cancer Informatics*, *4*, pp.637-646.

Meena Kowshalya, A. and Valarmathi, M.L., 2018. Evaluating twitter data to discover user's perception about social internet of things. *Wireless Personal Communications*, *101*, pp.649-659.

Minovski, D., Ogren, N., Ahlund, C. and Mitra, K., 2021. Throughput prediction using machine learning in lte and 5g networks. *IEEE Transactions on Mobile Computing*.

Nath, D., Singh, M., Sethia, D., Kalra, D. and Indu, S., 2020, March. A comparative study of subject-dependent and subject-independent strategies for EEG-based emotion recognition using LSTM network. In *Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis* (pp. 142-147).

Nayak, R.K., Mishra, D. and Rath, A.K., 2015. A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. *Applied Soft Computing*, *35*, pp.670-680.

Nursal, a.t., improving residential housing project purchase by using integrated multi-attribute decision making and sentiment analysis technique.

Pang, B., Lee, L. and Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Prados-Peña, M.B., &Saez-Pérez, M.P. (2020, May 17). The conservation and restoration of heritage as an improvement in the competitiveness of tourist destinations through the vision of tourists. Case study of the Monumental Complex of the Alhambra and the Generalife. https://doi.org/10.5281/zenodo.6239363

Sarkodie, S.A. and Owusu, P.A., 2021. Global effect of city-to-city air pollution, health conditions, climatic & socio-economic factors on COVID-19 pandemic. *Science of the Total Environment*, *778*, p.146394.

Shaeeali, N.S., Mohamed, A. and Mutalib, S., 2020. Customer reviews analytics on food delivery services in social media: A review. *IAES Int. J. Artif. Intell*, *9*(4), pp.691-699.

Shamshirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A.T. and Alinejad-Rokny, H., 2021. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. *Journal of Biomedical Informatics*, *113*, p.103627.

Tang, D., Qin, B. and Liu, T., 2016. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*.

Vasconcelos, H., Jörke, M., Grunde-McLaughlin, M., Gerstenberg, T., Bernstein, M. and Krishna, R., 2022. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *arXiv preprint arXiv:2212.06823*.

Wu, Y., Chen, Z., Wu, L., Lin, P., Cheng, S. and Lu, P., 2017. An intelligent fault diagnosis approach for PV array based on SA-RBF kernel extreme learning machine. *Energy Procedia*, *105*, pp.1070-1076.