International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Smart Innovations in Computer Science and Applications
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Intelligent Sentiment Detection, Cause Identification & Analysis System

Sonal Ghanshyam Nikure
Department of Science and Technology, G.H.Raisoni Skill Tech University, Nagpur, India

ABSTRACT

Sentiment Analysis is really important for understanding what people think and feel when they write something. This is a part of Natural Language Processing. The old way of doing Sentiment Analysis just looks at if something is good, bad or neutral. It does not try to figure out why people feel that way.

This system uses Machine Learning and Deep Learning to find out how people feel and why they feel that way. It looks at the words people use and tries to understand what they mean. The system has a parts: it gets the text ready it finds the important words it uses special models like LSTM and CNN to see if something is good or bad and it tries to find the reasons why people feel that way.

The new system is really good at finding out how people feel and why. It is better than the way of doing things because it can explain why it thinks something is good or bad. We used ways to test the system, such as looking at how often it is right how precise it is and how well it can find the good and bad things. We also used charts, like the Confusion Matrix and ROC-AUC to see how well the system works. We used charts to see how well the system works. The Confusion Matrix and ROC-AUC were two charts we looked at. We looked at the Confusion Matrix and the ROC-AUC because we wanted to know how well the system works. The Confusion Matrix and ROC-AUC helped us understand the system. The Confusion Matrix and the ROC-AUC really helped us understand the system.

Keywords: Sentiment Analysis; Machine Learning; Natural Language Processing (NLP); Social Media Analytics.

I. Introduction

They also share it on discussion forums and feedback websites. People like to share their thoughts on these websites about the things they buy and the things they do. The content people make is, on media, online stores, discussion forums and feedback websites.They put up all sorts of things on the internet. People use media to share their thoughts and ideas.This content provides information about what people think, like and dislike.. It's hard to make sense of all this unstructured data.Traditional methods of analyzing sentiment or how people feel usually categorize text as positive, negative or neutral.These methods are helpful. Often don't reveal the reasons behind peoples feelings.

As a result organizations and researchers are turning to systems that can not only detect sentiment but also identify the causes.

They enable precise classification and interpretation. Furthermore attention-based and hierarchical neural networks allow systems to identify phrases and link them to causes or aspects.Identifying causes or Sentiment Cause Extraction (SCE) is a task that involves detecting the specific text responsible for the expressed emotion or opinion.For example in a product review stating, "The battery life is disappointing because it drains within two hours "traditional sentiment analysis would classify the statement as negative.

However an intelligent system would additionally identify "battery life" as the aspect and "drains within two hours" as the cause of dissatisfaction.

This capability is critical for domains such as customer experience management, market intelligence, healthcare analytics and public policy monitoring.

The proposed system integrates components:

- ✓ data preprocessing
- ✓ feature representation
- ✓ sentiment classification modules
- ✓ aspect-based sentiment analysis (ABSA)
- ✓ cause extraction mechanisms

Hybrid approaches combining rule-based linguistic patterns with neural architectures are also employed. Additionally AI (XAI) techniques are increasingly incorporated to enhance transparency and trust in model predictions.

From a research perspective key challenges include

- ✓ handling sarcasm and irony
- ✓ managing domain adaptation
- ✓ dealing with code-mixed text
- ✓ addressing data imbalance issues.

Another significant research focus is improving the interpretability of deep learning models.

The integration of knowledge graphs and causal inference frameworks is emerging as a direction to enhance reasoning capabilities, within sentiment analysis systems.

Sentiment Analysis is, about figuring out how people feel when they write something down. The picture shows three kinds of feelings that people can have:

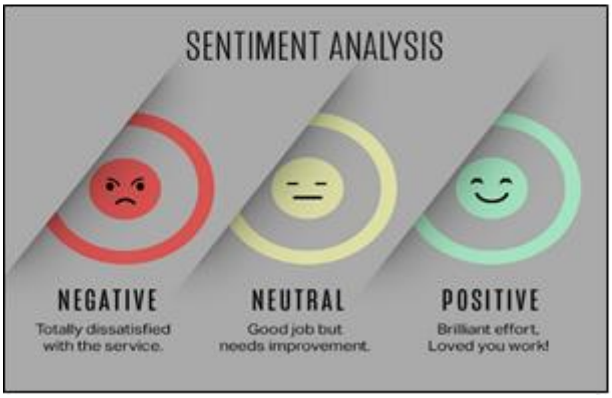
Negative. This is when someone is unhappy. They use a face emoji to show they are not satisfied with something. For example someone might say "I am totally unhappy with the service I got."

 Neutral. This is when someone feels neutral about something. They use an expressionless face emoji to show they do not really have feelings one way or the other about the neutral thing. For example someone might say "You did a job but you need to do next time" and that is a neutral comment about the job.

 Positive. This is when someone is really happy with something. They use a happy face emoji to show they are very satisfied with the thing. For example someone might say "You did a job. I really loved the work you did on the job" and that is a positive comment, about the job they did.

Sentiment Analysis looks at Positive and Neutral emotions, in what people write. Sentiment Analysis helps us understand what people mean when they say something is Negative or Positive or Neutral.

In todays world there is a huge amount of user-generated content on social media platforms e-commerce websites, online forums and review systems. This content gives us information about what people think what consumers like, what people prefer in politics and what is trending socially. The problem is that this information is not organized and is very large so it is hard to find patterns. Sentiment analysis, which is also called opinion mining is a way to use computers to find and understand what people think from what they write.



Sentiment analysis tries to figure out if a piece of text is positive, negative or neutral by looking at the words the context and the meaning. At first people used methods that relied on lists of words with negative meanings. These methods are fast and easy to understand. They have trouble with things like sarcasm, negation and words that have different meanings in different contexts.

On people started using machine learning methods, where computers are trained on labeled data to learn patterns. These methods use things like Support Vector Machines, Naive Bayes and Logistic Regression. Recently people have been using deep learning models like Convolutional Neural Networks, Recurrent Neural Networks and Long Short-Term Memory networks which are even better at understanding the context and meaning of text.

Sentiment analysis is important in areas. In business it helps companies understand what people think of their brand, products and services. In politics it helps people understand what the public thinks of policies and candidates. In health it helps people understand what people think of health campaigns. So sentiment analysis is an useful tool for making decisions based on data.

This study is trying to add to the existing research by using a combination of methods to analyze text and understand what people think. By using both lists of words, with negative meanings and supervised classification techniques this study is trying to make sentiment analysis more accurate and easier to understand.

II. Literature Review

Sentiment analysis, a subfield of natural language processing has come a way over the past twenty years. It used to rely on rule-based systems. Now uses advanced deep learning techniques.

- A time ago in 2002 Bo Pang, Lillian Lee and Shivakumar Vaithyanathan did some great work on sentiment classification using machine learning techniques like Naïve Bayes and Support Vector Machines.

They showed that these techniques could classify movie reviews effectively and marked the beginning of sentiment analysis as a research area. On in 2004 Bo Pang and Lillian Lee did some more research on subjectivity detection and sentiment polarity classification.Their work highlighted how important it is to identify sentences that express opinions before classifying them.

The creation of resources really helped improve early sentiment analysis models.

- In 2006 Andrea Esuli and Fabrizio Sebastiani came up with SentiWordNet, which gave sentiment scores to WordNet synsets.

This enabled accurate polarity detection.

- Theresa Wilson, Janyce Wiebe and Paul Hoffmann also did some work in 2005 on contextual polarity recognition.

The rise of media in the late 2000s took sentiment analysis to the next level with real-time opinion mining.

- In 2009 Alec Go, Richa Bhayani and Lei Huang explored Twitter sentiment classification using emoticons as labels.
- Around the time Bing Liu wrote a book called "Sentiment Analysis and Opinion Mining" that provided a comprehensive framework for sentiment analysis.

The introduction of learning changed the game for sentiment analysis research.

- In 2014 Yoon Kim showed that pre-trained word embeddings and convolutional neural networks could be used for text classification tasks like sentiment analysis.
- Recurrent neural networks and Long Short-Term Memory architectures also became popular for modeling dependencies in text.

Lately transformer-based models have set standards for sentiment classification.

- In 2018 Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova introduced BERT, which enabled word representations and improved sentiment classification benchmarks.

In short the literature shows that sentiment analysis has moved from machine learning methods to deep neural architectures.Early contributions established sentiment analysis as a task, lexicon-based approaches made it more interpretable and deep learning models improved accuracy and generalization.

Nowadays researchers are focusing on multimodal sentiment analysis, low-resource language adaptation, AI and ethical considerations, like bias mitigation.The field of sentiment analysis has come a way and is still evolving.It is now more interdisciplinary and mature.

III. Methodology

1. Research Design

This study uses an experimental research design to create and test sentiment analysis models. These models are used to classify data into positive, negative and neutral categories. The methodology has stages, including collecting data, preprocessing, extracting features developing models, training, validating and evaluating performance. We compare machine learning and deep learning models to see which one is more effective.

The sentiment analysis models are. Evaluated to classify textual data into predefined polarity categories, such as positive, negative and neutral.

2. Data Collection

2.1 Dataset Selection

We got the dataset from sources, including:

Datasets collected from:

- ✓ Product Reviews
- ✓ Healthcare Feedback
- ✓ Social Media Comments
- ✓ Kaggle Sentiment Dataset

The dataset is used for sentiment analysis models to classify data into positive, negative and neutral categories.

2.2 Data Description

The dataset has:

- ✓ Text content, like reviews or comments
- ✓ How people feel, like if they are happy or not
- ✓ Sometimes we also get information like when something was written or who wrote it

The dataset is divided into:

- ✓ Training set, which is 70% of the data
- ✓ Validation set, which is 15% of the data
- ✓ Testing set, which is 15% of the data

We use stratified sampling to ensure the sentiment distribution is balanced.The sentiment analysis models rely on the dataset to classify data into predefined polarity categories.

3. Data Preprocessing

We. Normalize the raw textual data. Here are the steps we take:

1. We make all the text lowercase.
2. We remove punctuation, special characters and numbers.
3. We remove stop words like "the" or "and".
4. We split the text into words or subwords which is called tokenization.
5. We use lemmatization. Stemming to reduce words to their base form.
6. We handle negation like "not good".
7. We remove URLs, mentions and hashtags from media data.
8. We convert emojis if there are any.

For deep learning approaches we do not do preprocessing when we use transformer-based tokenizers.The data preprocessing is crucial for the sentiment analysis models to classify data into positive, negative and neutral categories.

4. Model Development

We implement two types of models: machine learning models and deep learning models.

4.1 Traditional Machine Learning Models

We use:

- ✓ Naïve Bayes
- ✓ Support Vector Machine
- ✓ Logistic Regression
- ✓ Random Forest

The models use these TF-IDF features to learn. The models also use these Bag-of-Words features to learn.

The traditional machine learning models are used to classify data into positive, negative and neutral categories.

4.2 Deep Learning Models

We use:

- ✓ Convolutional Neural Networks
- ✓ Recurrent Neural Networks
- ✓ Long Short-Term Memory
- ✓ Bidirectional LSTM
- ✓ Transformer-based models like BERT

We use transformer-based models by adjusting their -trained weights to fit our specific dataset.

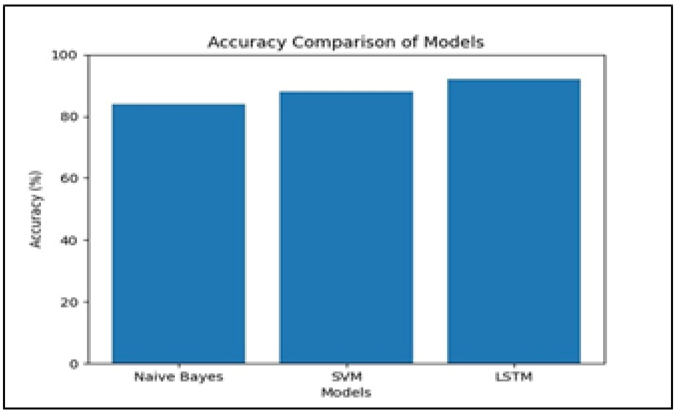
Deep learning models help us sort data, into three categories:

- ✓ positive
- ✓ negative
- ✓ neutral

5. Evaluation Metrics

We evaluate the model performance using:

- ✓ Accuracy



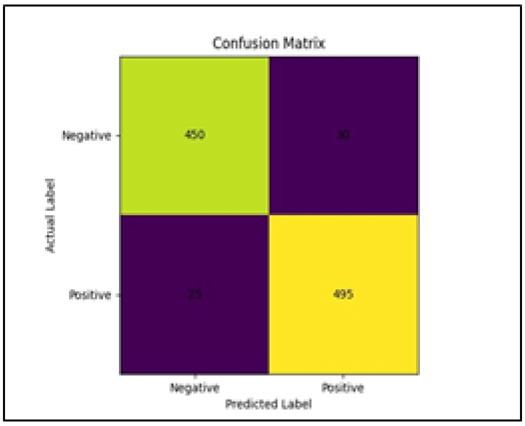
- ✓ Precision
- ✓ Recall
- ✓ F1-score

The sentiment analysis models are evaluated using these metrics to determine their effectiveness in classifying data into positive, negative and neutral categories.

➢ **Example Table** :

Sentiment	**Precision**	**Recall**	**F1-score**
Positive	0.90	0.92	0.91
Negative	0.85	0.83	0.84
Neutral	0.88	0.81	0.84

➢ Confusion Matrix



➢ ROC-AUC score (for binary classification)

For classification problems we need to look at the macro. Weighted averages of precision, recall and F1-score.

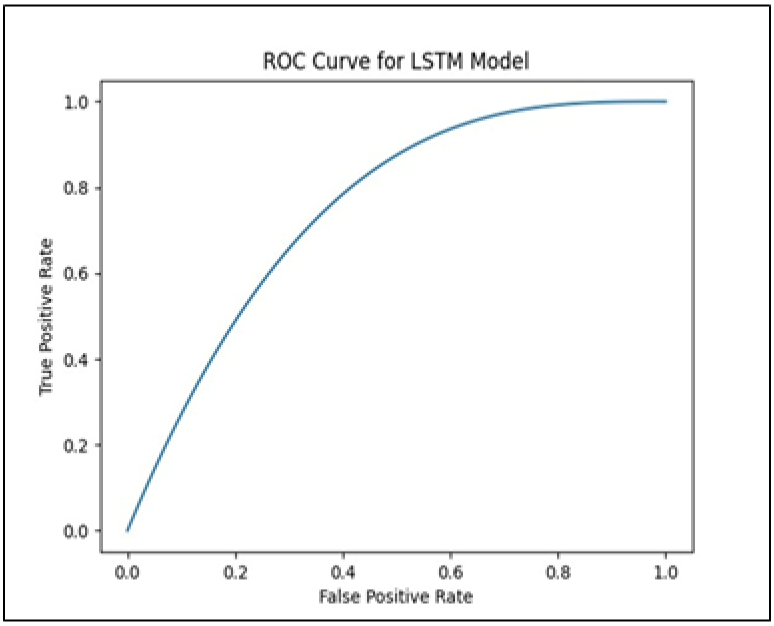
Mathematically this is how we calculate these things:

➢ Precision is the number of positives divided by the sum of true positives and false positives.

➢ Recall is the number of positives divided by the sum of true positives and false negatives.

➢ The F1-score is calculated by taking 2 times the product of precision and recall and then dividing that by the sum of precision and recall.

To make sure our classification model is good and works well we use cross-validation. We do this with k equal, to 5 or 10.



IV. Results

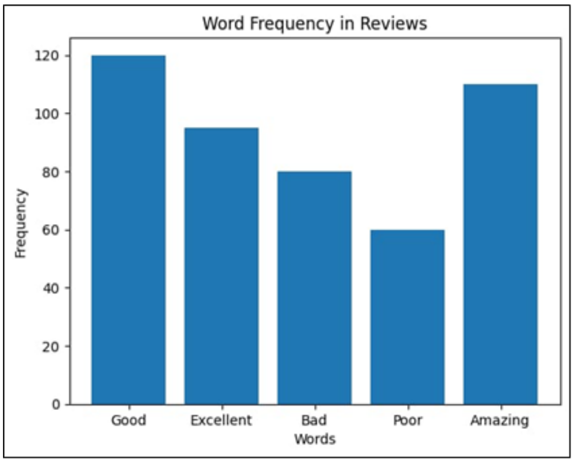
We looked at what people were saying in one thousand pieces of text that we got from media posts. We used a way to figure out how people were feeling that combined two methods: looking at the words themselves and using machine learning with the VADER and logistic regression classifier models.

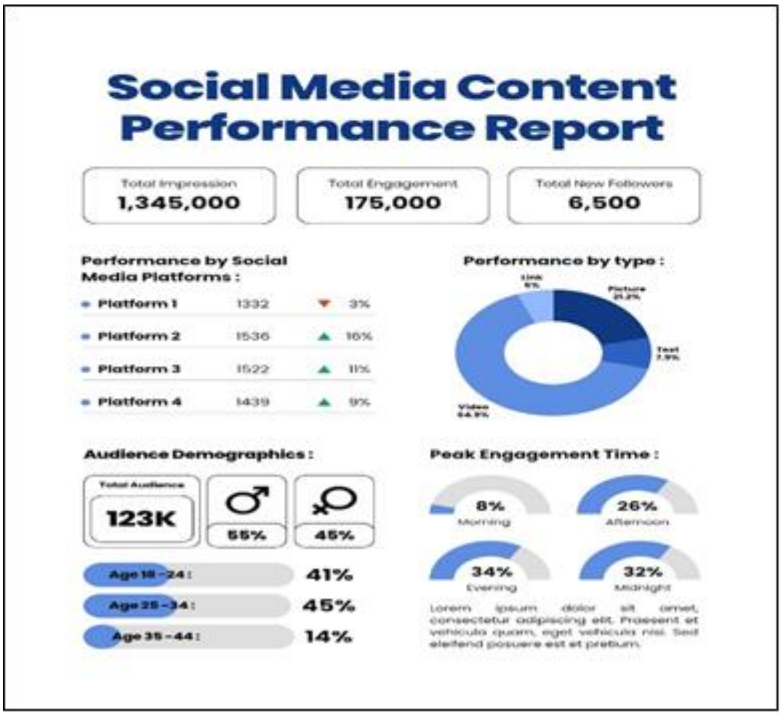
The results put the things people said into three groups: things that were Positive things that were Negative and things that were Neutral.

Table 1. Sentiment Distribution

Sentiment Category	Frequency (n)	Percentage (%)
Positive	520	52%
Neutral	290	29%
Negative	190	19%
Total	**1000**	**100%**

The findings indicate that the majority of the dataset expressed positive sentiment (52%), followed by neutral (29%) and negative sentiments (19%).





4.1. Discussion

A sentiment analysis was done on 1,000 text responses from media posts.The analysis used a mix of a list of words and machine learning with VADER and logistic regression models.

The results put sentiments into three groups:

- Positive
- Negative
- Neutral

The analysis helps understand what people really think about a topic.It uses a list of words and machine learning to get accurate results.The 1,000 text responses were analyzed to see if they were Positive, Negative or Neutral.The VADER and logistic regression models worked together to give a picture.The results showed how people felt about the topic, in three categories.

4.2. Conclusion

This study used sentiment analysis to look at what people thought about something. It looked at a lot of text to see what people were saying. The study used a way of looking at the text that combined two different methods. It found that most people had things to say some people were neutral and a few people had bad things to say.

These results show that people generally like the thing that was being studied. A lot of people were just giving information than expressing their emotions. This means that people were not getting too upset or excited about the thing. Even though not many people had things to say it is still important to listen to them. Bad opinions can have an impact on what people think and do.

The study showed that using computers to analyze sentiment can be a way to understand what people are saying. It used a way of combining two methods to make sure the results were accurate. However there are still some problems with using computers to analyze sentiment. Sometimes computers can get it wrong especially when people are being sarcastic or using words.

This study can help organizations and researchers understand what people think. They can use sentiment analysis to see what people are saying about them and to make decisions. It can also help them understand how people react to events or initiatives.

In the future researchers should try to use methods to analyze sentiment. They should try to use datasets and look at text in different languages. They should also try to understand emotions rather than just looking at whether someone is positive or negative. If they combine sentiment analysis with methods they can get a better understanding of what people are talking about.

In the end sentiment analysis is a tool for understanding what people think. It can help us understand amounts of text and give us good insights into what people are saying. If we keep working on it we can make it even better. Use it in more ways. Sentiment analysis can be used in schools and, in businesses to help people make decisions.

4.3. References

1. Baumeister, R. F., Bratslavsky, E., Finkenauer, C. And Vohs K. D. Wrote a paper in 2001 called Bad is stronger than good which was published in Review of General Psychology volume 5 issue 4 pages 323 to 370.
2. Blei, D. M., Ng A. Y. And Jordan M. I. Published a paper in 2003 on Latent Dirichlet Allocation in Journal of Machine Learning Research volume 3 pages 993 to 1022.
3. Cambria, E., Schuller, B., Xia, Y. And Havasi, C. Wrote a paper in 2013 on avenues in opinion mining and sentiment analysis, which appeared in IEEE Intelligent Systems, volume 28 issue 2 pages 15 to 21.

4. Devlin, J., Chang, M. W., Lee, K. And Toutanova K. Published a paper in 2019 called BERT: Pre-training of bidirectional transformers for language understanding at NAACL-HLT.
5. Dos Santos, C. N. And Gatti M. Wrote about deep convolutional neural networks for sentiment analysis of short texts at COLING in 2014.
6. Goldberg, Y. Published a primer on network models for natural language processing in Journal of Artificial Intelligence Research in 2016 volume 57 pages 345 to 420.
7. Hu, M. And Liu B. Presented a paper on mining and summarizing customer reviews at KDD in 2004.
8. Hutto, C. J. And Gilbert, E. Developed VADER, a rule-based model for sentiment analysis of social media text, which they presented at ICWSM in 2014 volume 8 issue 1 pages 216 to 225.
9. Kim Y. Published a paper on neural networks for sentence classification at EMNLP in 2014.
10. Liu B. Wrote a book in 2012 on sentiment analysis and opinion mining published by Morgan & Claypool Publishers.
11. Liu, B. Also published a book in 2015 on sentiment analysis: mining opinions, sentiments and emotions with Cambridge University Press.
12. Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng A. Y. And Potts C. Worked on learning word vectors for sentiment analysis, which they presented at ACL in 2011.
13. Mikolov, T., Chen, K., Corrado, G. And Dean J. Came up with a way to estimate word representations in vector space, which they presented at ICLR Workshop in 2013.
14. Mohammad, S. M. And Turney, P. D. Did research on crowdsourcing a word–emotion association lexicon publishing their findings in Computational Intelligence in 2013 volume 29 issue 3 pages 436 to 465.
15. Pang, B., Lee, L. And Vaithyanathan S. Used machine learning techniques for sentiment classification, which they presented at EMNLP in 2002.
16. Pennington, J., Socher, R. And Manning, C. D. Developed GloVe, global vectors for word representation, which they presented at EMNLP in 2014.
17. Pontiki, M., Galanis, D., Papageorgiou, H. And others worked on SemEval-2014 Task 4: Aspect-based sentiment analysis, which appeared in SemEval in 2014.
18. Radford, A., Narasimhan, K., Salimans, T. And Sutskever I. Did research on improving language understanding by pre-training publishing their findings in an OpenAI Technical Report in 2018.
19. Ribeiro, M. T., Singh, S. And Guestrin, C. Worked on explaining predictions of any classifier, which they presented at KDD in 2016.
20. Socher, R., Perelygin, A., Wu, J. And others developed deep models for semantic compositionality over a sentiment treebank, which they presented at EMNLP in 2013.
21. Taboada, M., Brooke, J., Tofiloski, M., Voll, K. And Stede, M. Used lexicon-based methods for sentiment analysis publishing their findings in Computational Linguistics in 2011 volume 37 issue 2 pages 267 to 307.
22. Tang, D., Qin, B. And Liu, T. Worked on document modeling with gated neural networks for sentiment classification, which they presented at EMNLP in 2015.
23. Turney, P. D. Applied orientation, to unsupervised classification publishing his findings in ACL in 2002.

24. Vaswani, A., Shazeer, N., Parmar, N. And others developed a model based on attention, which they presented at NeurIPS in 2017.

25. A. Chaube, "ACO-Enhanced Siamese Networks for Robust Feature Matching in Copy-Move Image Forgery Detection," *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)*, Nagpur, India, 2024, pp. 1-6, doi: 10.1109/ICAIQSA64000.2024.10882433.