Int. J. Sci. R. Tech., 2025 2(3)

A Multidisciplinary peer-reviewed Journal www.ijsrtjournal.com [ISSN: 2394-7063]

Sentimental Analysis on Veganism

Omkar Singh¹, Sunali Bhattacherji^{*2}

¹HOD (Department of Data Science), Thakur College of Science & Commerce ²PG Student, Department of Data Science, Thakur College of Science & Commerce Thakur Village, Kandivali (East), Mumbai-400101, Maharashtra, India

ABSTRACT

This project introduces a sentiment analysis system developed to gauge public opinions on veganism as a social justice movement. By employing machine learning algorithms—Support Vector Classification (SVC), Logistic Regression, and K-Nearest Neighbors (KNN)—the system categorizes text data into positive, negative, or neutral sentiments. A dataset of 50,000 text entries from Kaggle was preprocessed and converted using Term Frequency-Inverse Document Frequency (TF-IDF). Comparative analysis using accuracy, precision, recall, and F1-score determined the most effective model. The scalable system supports social media monitoring and public perception research, providing insights into societal views on veganism.

Keywords: Sentiment Analysis, Veganism, Social Justice, Machine Learning, Support Vector Classification (SVC), Logistic Regression, K-Nearest Neighbors (KNN), Term Frequency-Inverse Document Frequency (TF-IDF), Public Perception, Animal Rights

INTRODUCTION

Veganism is a social justice movement dedicated to ending the exploitation and oppression of animals. Advocates view it as an ethical commitment to stop using animals as commodities, emphasizing their right to live without harm. The movement challenges long-standing norms that have accepted animal exploitation, promoting both individual responsibility and broader societal change. As veganism gains traction, it elicits a range of public responses, influencing societal acceptance and policy decisions. This project aims to develop a sentiment analysis system to gauge public attitudes toward veganism, providing valuable insights for advocacy and fostering informed discussions about animal rights. By examining language and tone in online conversations, the system can identify patterns in public perception and track changes in understanding and support. This detailed analysis can help vegan advocates and animal rights organizations tailor their strategies, contributing to a more empathetic public awareness of animal rights issues. Ultimately, the tool seeks to support the vegan movement by clarifying perceptions and encouraging public ethical discussions about animal treatment. Additionally, understanding these sentiments can highlight areas of resistance and support, guiding more effective outreach and education efforts. By leveraging machine learning, this project aims to offer a scalable and robust solution for monitoring public opinion on veganism, aiding in the broader goal of achieving justice and compassion for all sentient beings.

LITERATURE REVIEW:

Rikters, M., & Kale, M. (2021) analyze Twitter sentiments on meat consumption over ten years, emphasizing its health and environmental impacts. They use sentiment analysis to categorize tweets in Latvian, revealing public attitudes towards meat and alternative proteins. The study highlights the environmental cost of meat production and the influence of seasonal food preferences, advocating for interdisciplinary research to address food consumption and sustainability.

Shamoi et al. (2022) analyze public sentiment on vegan diets using Twitter data, employing mutual information for feature selection. The study reveals a positive shift in sentiment towards veganism over 12 years, driven by health benefits, COVID-19, and climate concerns. The authors emphasize the

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



importance of sentiment analysis in promoting healthy, sustainable eating habits.

Jennings et al. (2021) investigate perceptions of veganism through surveys and social media analysis. They find that non-vegans are skeptical about the health benefits of veganism and perceive it as less healthy and difficult compared to vegans. Social media analysis reveals positive sentiment towards veganism, suggesting that leveraging social media could help improve public perceptions and encourage adoption. The study highlights the role of social media in shaping dietary behaviors and calls for further research on its impact on public health initiatives.

Park and Kim (2022) investigate how vegans and nonvegans view veganism during the COVID-19 pandemic. Through Word2Vec and qualitative analysis of Reddit discussions, they explore key aspects of veganism, including lifestyle, animal rights, and food. The study reveals biases against veganism among nonvegans and examines how the pandemic has influenced food choices. The authors propose that understanding these differing perspectives can help address biases and encourage the adoption of veganism.

Kadel et al. (2024) explore how Instagram influences perceptions of veganism and its effect on eating intentions. By analyzing 44,316 posts tagged with #vegan, they discover that content frequently focuses on food, health, cosmetics, and photography, with a generally positive sentiment. The study indicates that viewing vegan content on Instagram is associated with eating intentions, with attitude and self-identity playing significant roles. The authors emphasize the potential of social media to encourage healthy eating habits and suggest further research on its impact on dietary choices.

Gangrade, Shrivastava, and Gangrade (2018-19) investigate sentiment analysis on Instagram using natural language processing and Thayer's psychologically defined model. They propose a method for classifying sentiments by extracting keywords from hashtags, achieving an accuracy rate of 90.7%. The study highlights Instagram's role in emotional expression and suggests that their method offers a nuanced understanding of user emotions, surpassing traditional polarity classification. The authors recommend applying this approach to analyze social phenomena across various fields.

Karimvand et al. (2018-19) introduce a multimodal deep learning method for sentiment analysis of Persian Instagram posts, using a bi-directional gated recurrent unit for text and a 2-dimensional convolutional neural network for images. Their new dataset, MPerInst, demonstrates that combining text and image modalities significantly enhances sentiment detection accuracy and F1-score. The proposed model outperforms existing deep fusion models and highlights the potential of multimodal approaches for analyzing social media sentiment. The authors advocate for further research in multimodal sentiment analysis and its diverse applications.

Architecture and Design:

a. Objective Definition:

The main goal is to create a dependable and precise machine learning-based sentiment analysis system that categorizes sentiments about veganism as positive, negative, or neutral. This system aims to analyze public opinion, offering valuable insights for advocates, researchers, and organizations dedicated to veganism as a social justice movement.

b. System Components:

1. Data Collection:

Gather a comprehensive dataset containing text data reflecting sentiments toward veganism. Ensuring a balanced dataset is crucial to accurately represent positive, negative, and neutral sentiments without bias. For this project, data has been sourced from Kaggle, ensuring a diverse range of opinions and contexts.

2. Data Preprocessing:

Conduct standard text cleaning and preprocessing to prepare the data for analysis. This involves converting text to lowercase, removing punctuation, and using TF-IDF vectorization to transform text into numerical features. Preprocessing is vital to ensure consistency and enhance the model's ability to generalize across various text sources.

3. Model Development:

Develop and evaluate four machine learning models—Support Vector Classification (SVC), Logistic Regression, and K-Nearest Neighbors (KNN)—to compare their performance. These models are selected for their varied strengths in handling different aspects of text classification, from boundary definition to similarity-based and ensemble learning approaches.

c. Implementation Plan:

Dataset Preparation: The dataset, sourced from Kaggle, consists of 50,000 rows of text data labeled with positive, negative, or neutral sentiments toward veganism. Preprocessing involves standardizing the input by converting text to lowercase and removing punctuation or non-alphabetic characters. To enhance the model's generalization capabilities, 10% of sentiment labels are modified to simulate noise. Text data is then transformed into numerical features using TF-IDF vectorization with n-grams (1-2) and a feature limit of 5,000 to capture key patterns and phrases.

Model Architecture: The proposed system architecture integrates multiple components to process and analyze sentiment data effectively.

Feature Extraction: To optimize text data for machine learning, TF-IDF vectorization is applied. This method transforms text data into numerical features using Term Frequency-Inverse Document Frequency (TF-IDF) with n-grams (1-2) and a maximum of 5,000 to 6,000 features, capturing nuanced patterns and phrases to enhance the model's ability to differentiate sentiment classes effectively.

Model Implementation and Training: Three machine learning models are implemented and trained: Support Vector Classification (SVC) for its precise boundary definition, Logistic Regression as a baseline model for its interpretability, and K-Nearest Neighbors (KNN) for its similarity-based classification to capture sentiment trends in context.

Training and Validation: The dataset is split into training (80%) and validation (20%) sets. k-Fold Cross-Validation is utilized to enhance model generalization and monitor for overfitting. Performance metrics such as accuracy, precision, recall, and F1 score are tracked to optimize model performance.

Model Evaluation: Overall accuracy is measured to understand the proportion of correct predictions across all sentiment classes. A detailed classification report is generated to review precision, recall, and F1score for each class (positive, negative, neutral), providing insights into the model's performance on individual sentiments. The weighted F1 score is calculated to balance precision and recall, especially useful for assessing performance across imbalanced sentiment classes. The confusion matrix is analyzed to observe the distribution of true positives, true negatives, false positives, and false negatives, helping identify patterns in misclassification and areas where the model may need refinement.

Deployment Strategy: The model is deployed on a cloud platform to provide scalable, remote access for social justice organizations and researchers.

Training and Support: User training materials are developed to help users effectively monitor and analyze public opinion. Ongoing technical support is offered to ensure effective implementation and troubleshooting.

Feedback Loop: A user feedback mechanism is established for continuous improvement based on user feedback. The model is regularly updated with new data to maintain accuracy and relevance.

Diagram for data flow diagram of methodology:



1. Dataset Collection

The research starts with gathering a comprehensive dataset of 50,000 text samples related to veganism. This dataset is categorized into three sentiment classes: positive, negative, and neutral. Sourced from Kaggle, it includes data from various online platforms, including social media, ensuring it reflects real-world sentiments and opinions about veganism.

2. Data Preprocessing

To enhance the model's performance and ensure consistency in the input data, extensive preprocessing is conducted. This includes:

Text Cleaning: Converting all text to lowercase and removing non-alphabetic characters using regular expressions. Noise Simulation: Introducing noise by randomly altering 5-10% of sentiment labels to reduce predictability and improve generalization.

TF-IDF Vectorization: Transforming the cleaned text into numerical features using the TF-IDF method, with a feature limit of 6,000 to optimize model performance.

3. Dataset Splitting

The dataset is systematically divided into three subsets: training, validation, and test sets, using an 80-10-10 ratio. This strategic division allows for effective model training, hyperparameter tuning, and final evaluation, while maintaining class distribution through stratified sampling.

4. Model Development



The project implements three distinct models for sentiment classification:

- 1. Support Vector Machine (SVM): Employed for its effectiveness in high-dimensional spaces, using a linear kernel to classify sentiments based on features extracted from the TF-IDF representation.
- 2. Logistic Regression: Trained with class weights to manage class imbalance, enhancing its robustness in sentiment classification.
- **3. K-Nearest Neighbors (KNN):** Utilizes distance metrics to classify sentiments based on proximity to labeled data points in feature space.

5. Hyperparameter Tuning

Hyperparameter optimization is performed using GridSearchCV across all models to maximize their performance. This involves systematic exploration of key parameters, such as regularization strength for SVM, maximum iterations for logistic regression, and the number of neighbors and weight function for KNN.

6. Model Evaluation

The performance of each model is rigorously evaluated using metrics including accuracy, precision, recall, F1-score, and ROC-AUC. The evaluation is conducted on the test set, which remains untouched during training, ensuring the validity of results.

7. Analysis and Interpretation

Following evaluation, a detailed analysis of each model's performance is conducted. The results are examined for balance across classes, ensuring no class is favored. Insights drawn from precision and recall metrics are crucial in determining the models' effectiveness in capturing sentiments towards veganism.

8. Deployment and Validation in Real-World Settings

The final models are deployed for practical testing on real-world text data related to veganism. User feedback and performance metrics from this deployment inform iterative improvements to enhance model reliability and usability in diverse contexts.

9. Conclusion and Future Work

The study concludes that the developed sentiment analysis system can assist in understanding public sentiments towards veganism, providing valuable insights for further research and applications in health and dietary fields. Future work will focus on expanding the dataset and exploring advanced machine learning techniques to improve model accuracy and robustness.

RESULTS AND DISCUSSION:

Upon evaluating the models on the test set, the following key metrics were obtained for each algorithm:

Support Vector Machine (SVM):

- 1. Support Vector Machine (SVM):
- o Best C Parameter: 0.001
- o **Accuracy:** 93.44%
- o **F1 Score:** 0.934

. .

. . .

- o Classification Report:
- Precision: 94.0% (Negative), 93.0% (Neutral), 93.0% (Positive)
- Recall: 93.0% (Negative), 94.0% (Neutral), 93.0% (Positive)
- Support: 3,332 (Negative), 3,331 (Neutral), 3,337 (Positive)

	o Confusion Matrix:		
	Predicted Negative	Predicted Neutral	Predicted Positive
Actual Negative	3112	109	111
Actual Neutral	101	3123	107
Actual Positive	115	113	3109

2. Logistic Regression:

- o Accuracy: 96.91%
- o Classification Report:

 Precision: 97.0% (Negative), 97.0% (Neutral), 97.0% (Positive)

- Recall: 97.0% (Negative), 97.0% (Neutral), 97.0% (Positive)
- Support: 3,336 (Negative), 3,327 (Neutral), 3,337 (Positive)
- 3. K-Nearest Neighbors (KNN):
- o Best Parameters: {'n_neighbors': 5, 'weights':
 'uniform'}
- o **Accuracy:** 93.44%

- o Classification Report:
- Precision: 94.0% (Negative), 93.0% (Neutral), 93.0% (Positive)
- Recall: 93.0% (Negative), 94.0% (Neutral), 93.0% (Positive)
- Support: 3,332 (Negative), 3,331 (Neutral), 3,337 (Positive)

Representation of Models Accuracy:



Graphical Representation of the dataset



Comparative Analysis of the Model



Comparison of Performance Metrics across Models

DISCUSSION:

The Logistic Regression model demonstrated superior performance compared to both the SVM and KNN classifiers, achieving an impressive accuracy rate of 96.91%. All models exhibited high precision, recall, and F1-scores, indicating their effectiveness in identifying sentiments related to veganism. Notably, the SVM model, despite having a slightly lower accuracy, excelled in recall for neutral sentiments, underscoring its reliability in distinguishing between different sentiment classes. The consistent performance across all models suggests that the preprocessing and feature extraction methods used were highly effective. Techniques such as tokenization, stop-word removal, and TF-IDF vectorization likely contributed to the models' ability to accurately capture and analyze sentiment nuances. Additionally, the balanced nature of the dataset ensured that the models could learn effectively from a diverse range of sentiment expressions, avoiding biases that could arise from an imbalanced dataset.

Furthermore, the robustness of the Logistic Regression model in this context can be attributed to its simplicity and efficiency in handling highdimensional data. The SVM model's strong performance in recall for neutral sentiments highlights its capability in handling cases where sentiment distinctions are subtle, which is crucial for comprehensive sentiment analysis. In conclusion, the experimental results validate the effectiveness of the implemented machine learning techniques for sentiment analysis in the context of veganism. Future research could explore more complex models, such as ensemble methods or deep learning approaches, to further enhance classification accuracy and address any misclassifications, particularly those involving sentiments with ambiguous or nuanced language. Additionally, incorporating advanced natural language processing techniques, such as word embeddings and contextualized language models like BERT, could provide deeper insights and improve the models' ability to understand and classify sentiments accurately.

CONCLUSION:

In this research, we developed an effective sentiment analysis model to classify sentiments related to veganism, utilizing a combination of machine learning techniques such as Support Vector Machines (SVM), Logistic Regression, and K-Nearest Neighbors (KNN). Through a thorough experimental setup, we assessed the performance of each model using a balanced dataset of 50,000 textual samples, ensuring a reliable evaluation of their effectiveness in distinguishing between negative, neutral, and positive sentiments. The findings revealed that the Logistic Regression model achieved the highest accuracy at 96.91%, demonstrating its ability to accurately classify sentiments with minimal errors. The SVM and KNN models also performed well, each achieving accuracies of 93.44%. The consistent precision, recall, and F1-scores across all models indicate their reliability and effectiveness in capturing the nuances sentiments of expressed about veganism. Furthermore, the study emphasized the importance of preprocessing and feature extraction in enhancing model performance. By employing techniques such as text normalization and TF-IDF vectorization, we ensured that the models could effectively learn from the dataset, leading to improved generalization on unseen data. These findings highlight the potential of machine learning techniques in sentiment analysis, particularly in addressing topics like veganism, which often evoke diverse opinions and emotional responses. Future research could focus on refining the models through advanced techniques, including deep learning and ensemble methods, as well as exploring larger and more diverse datasets to further enhance classification accuracy. Overall, this study contributes to the growing body of literature on sentiment analysis, providing a framework for future investigations into the complexities of public sentiments surrounding veganism and related dietary choices. The successful application of machine learning in this domain could lead to valuable insights for policymakers, marketers, and researchers interested in understanding consumer behavior and attitudes towards veganism.

REFERENCE

- Rikters, M., & Kāle, M. (2023). The Future of Meat: Sentiment Analysis of Food Tweets. In Proceedings of the 11th International Workshop on Natural Language Processing for Social Media (pp. 38-46).
- 2. Shamoi, E., Turdybay, A., Shamoi, P., Akhmetov, I., Jaxylykova, A., & Pak, A. (2022). Sentiment

analysis of vegan-related tweets using mutual information for feature selection. PeerJ Computer Science, 8, e1149.

- Jennings, L., Danforth, C. M., Dodds, P. S., Pinel, E., & Pope, L. (2019). Exploring perceptions of veganism. arXiv preprint arXiv:1907.12567.
- 4. Park, E., & Kim, S. B. (2022). Veganism during the COVID-19 pandemic: Vegans' and nonvegans' perspectives. Appetite, 175, 106082.
- Kadel, P., Heist, N., Paulheim, H., & Mata, J. (2024). From Pixels to Palate: Communication Around #vegan on Instagram and Its Relation with Eating Intentions. Appetite, 107518.
- Gangrade, S., Shrivastava, N., & Gangrade, J. (2019). Instagram sentiment analysis: opinion mining. Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA), April 16.
- Karimvand, A. N., Chegeni, R. S., Basiri, M. E., & Nemati, S. (2021). Sentiment analysis of Persian Instagram posts: a multimodal deep learning approach. In 2021 7th International Conference on Web Research (ICWR) (pp. 137-141). IEEE

HOW TO CITE: Omkar Singh, Sunali Bhattacherji *, Research on Sentimental Analysis on Veganism, Int. J. Sci. R. Tech., 2025, 2 (3), 450-457. https://doi.org/10.5281/zenodo.15082266