Catsonic: Cat Sound Classification Using Machine Learning Algorithms

Sarah D. Oghenekaro, Janelli M. Mendez, DIT

Lorma Colleges – College of Computer Studies and Engineering.
CLI Bldg. San Juan Campus, San Juan La Union
sarah.oghenekaro@lorma.edu, janelli.mendez@lorma.edu

Abstract

Catsonic is a unique project aimed at helping cat owners and veterinarians understand the sounds that cats make. Cats communicate through various sounds, and deciphering these can be challenging. In collaboration with cat owners and veterinarians, the researchers gathered a diverse set of cat sounds both online and from local communities in City of San Fernando, La Union. Cat sounds dataset forms the basis for training our machine learning model. The researchers approach involves using advanced technology called Convolutional Neural Networks (CNN) to classify and recognize different cat sounds. The researchers rigorously tested and evaluated the model to ensure accuracy and effectiveness in identifying various sounds, considering factors like sensitivity and specificity. The result of the researchers' efforts is a user friendly mobile application. Catsonic provides a tool for cat owners and veterinarians to interpret and understand cat sounds better. The researchers also assessed the application's ease of use through the System Usability Scale (SUS), the developed application was evaluated in terms of its usability and found to be acceptable with a percentage of 80.13%. Catsonic is not just an innovative solution; Catsonic serves as a bridge for improved communication between humans and their feline friends. This project showcases the power of technology, specifically machine learning, in addressing real challenges faced by pet owners and veterinarians. This capstone project highlights the transformative power of technology, specifically machine learning, in addressing tangible challenges faced by pet owners and veterinarians. Catsonic stands as a testament to the potential of leveraging advanced technologies for the betterment of the human-animal bond, showcasing how innovation can enhance understanding and connection with beloved pets.

Keywords**:** *Cat Sounds, Cat Owners, Cat Communication, Sound Classification, Machine Learning*

1. Introduction

Technology has been making life better for some time now, from banking, transportation, public administration, security and even entertainment, there are digital solutions. Machine learning, one of the numerous fields in which ICT can be applied, is the use and development of computer systems that can learn and adapt without following explicit instructions (Bostech ICT Solutions, 2019). Machine learning uses algorithms and statistical models to analyze and draw inferences from patterns in data. Machine learning is already being utilized in various software to create more efficient systems and enhance user experience. Machine learning is used to recognize and identify patterns, for instance with an e-commerce site, it can be used to monitor browsing patterns and past purchases.

Machine learning uses programmed algorithms that receive and analyze input data to predict output values within an acceptable range. As new data is fed to these algorithms, they learn and optimize their operations to improve performance, developing 'intelligence' over time (Wakefield, n.d). There are four types of machine learning algorithms: supervised, semi-supervised, unsupervised and reinforcement. Previous studies have explored the use of supervised learning algorithms for different areas of sound classification (Nanni 2020, Pandeya 2018, Chen 2019, Ferdiana 2021 etc.). In supervised learning example. The operator provides the machine learning algorithm with a known dataset that includes desired inputs and outputs, and the algorithm must find a method to determine how to arrive at those inputs and outputs. A running theme with these studies was the difficulty encountered when training the models due to limited data. Animal sound datasets are usually much smaller than necessary, since the sample collection and labelling can be expensive (Nanni, L., Maguolo, G., & Paci M, 2020). Commonly, audio signals can be augmented in the time and/or in the frequency domains directly on the raw signals or after their conversion into spectrograms (Nanni 2020). This challenge was overcome in Pandeya 2018 by using transfer learning in form a CNN pre trained for music recognition and data augmentation techniques.

The researchers project purposes to advance the development of applications that utilized precise sound recognition technologies which can be crucial in reducing or controlling problems like the rising threat of animal extinction or the effects of climate change on wildlife. The researchers aim is to utilize the findings of these studies to develop an application that can be useful to pet owners to better understand their feline friends. The application used a trained machine learning model to listen to cat sounds and infer what they mean. Benefits of using supervised machine learning algorithms to develop this application.

2. Objectives

This study developed an application that utilizes machine learning to accurately understand and interpret various sounds made by cats. The following statements are the specific objectives of Catsonic:

1. Collected and preprocessed the audio dataset for model training.
2. Trained the machine learning model for audio classification.
3. Evaluated the performance of the trained model in terms of; a. accuracy, b. sensitivity, and c. specificity.
4. Integrated the trained model into a functional mobile-based application; and

5. Determined the level of usability of the developed functional mobile-based application.

3. Materials and methods

The City of San Fernando, La Union served as the location for the pilot testing phase of the study, focusing on the development of a cat sound classification application. As the capital of the province of La Union, it was anticipated that the number of interested respondents, particularly cat owners and veterinarians, would be higher compared to other municipalities in the area. Moreover, the City of San Fernando hosts a substantial concentration of pet-related services, making it a suitable setting for the study. To select respondents, a simple random sampling technique was employed. This probability sampling method was instrumental in providing each member of the target population in La Union, comprising cat owners and veterinarians, with a fair and equal chance of being chosen. This approach facilitated an unbiased selection of respondents, contributing to the overall representativeness of the study's sample.

Ethical considerations were prioritized to ensure a comprehensive and diverse representation of respondents. Five (5) veterinarians contributed valuable insights into the medical and behavioral aspects of cat sounds, with their perspectives treated confidentially. Additionally, fifteen (15) cat pet owners, spanning various demographics, played a crucial role in providing real-world observations and experiences with their feline companions. The study's total of twenty (20) respondents aimed to strike a balance between veterinary professionals and cat owners, offering a holistic understanding of cat sounds from both medical and practical viewpoints.

The researchers assessed the performance of the trained model by employing the following metrics: accuracy, sensitivity, and specificity. These metrics provide a comprehensive understanding of the model's ability to correctly classify instances.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Sensitivity} = TP / (TP + FN)$$

$$\text{Specificity} = TN / (TN + FP)$$

In our pursuit of the fifth objective, the researchers chose to implement the System Usability Scale (SUS) as a tool to evaluate the user experience of the studied system. This approach involved utilizing a 10- item Likert scale questionnaire, where respondents, guided by the researcher's methodology, expressed their opinions on the ease of use for various technological applications. The aggregated responses provided valuable insights into the overall usability of the system, contributing to a comprehensive understanding of its user-friendliness.

The researchers applied the Cross Industry Standard Process for Data Mining (CRISP-DM) model, a well-established framework for guiding data mining endeavors. This model outlines the typical phases of a project, detailing the tasks within each phase and elucidating the relationships between these tasks (IBM Corporation, 2021). CRISP-

DM consists of six sequential phases, namely: business understanding, data understanding, data preparation, modeling, evaluation, and deployment, providing a structured representation of the data science life cycle as shown in Figure 1.

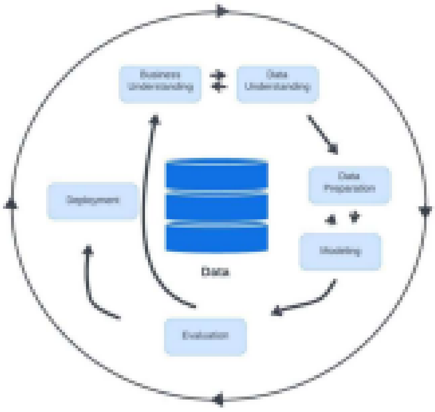

Figure 1. CRISP-DM Process Diagram

In the context of binary classification, the Confusion Matrix comprises four essential values: True Positive (TP), indicating the number of accurate positive predictions; True Negative (TN), denoting the count of correct negative predictions; False Positive (FP), representing the quantity of incorrect positive predictions; and False Negative (FN), signifying the tally of incorrect negative predictions.

Figure 2. Confusion Matrix for Binary Classification

4. Results

Understanding the various sounds cats make can be tricky for both cat owners and veterinarians. It’s important to notice the differences between meows, purrs, and hisses to communicate well with cat. To identify the problems that needed to be addressed relative to the problem statement deliverable, the researchers conducted an interview. A duly signed communication letter stating the said purpose was presented to the respondents. After talking to cat owners and veterinarians, it's clear that there's a gap in communication. This difficulty can make it hard to form strong connections between cats and their owners and might affect the health and happiness of their furry friends.

The identification and selection of appropriate data sources were essential for subsequent phases of the data mining process. Upon identifying the underlying problems in the respondents' cat sounds, the researchers identified several data sources relevant to the study. These include data from the respondents, audio recordings on cat sounds which the researchers gathered in person, and public datasets downloaded from the internet.

For preprocessing, the collected data was loaded into the researcher's development environment and visualized using the python libraries: librosa and matplotlib. This allowed the researchers to perform tasks like noise removal and feature extraction needed for the model training. As part of the data preprocessing phase, the data was renamed and organized into a data frame for model creation.

Table 1. Annotation after Data Preprocessing

New Tag	Classification
HA-0	Happy
SC-1	Scared
HU-2	Hungry
UL-3	Unlabelled

In the construction of the CNN model, three convolutional layers are employed. These layers have filter sizes of (3, 3) and increasing depths of 32, 64, and 128, respectively. After each convolutional layer, max-pooling layers with (2, 2) pool sizes are incorporated to down sample spatial dimensions. This design choice is influenced by the characteristics of the mel-frequency spectrogram data, with the aim of capturing hierarchical features at varying levels. Following the convolutional layers, a dense layer with 128 neurons and Rectified Linear Unit (ReLU) activation function is added. To mitigate overfitting, a dropout layer with a dropout rate of 30% is introduced before the final output layer. The output layer, responsible for classification, consists of three neurons representing the classes in the dataset. The softmax activation function is applied for multi-class classification.

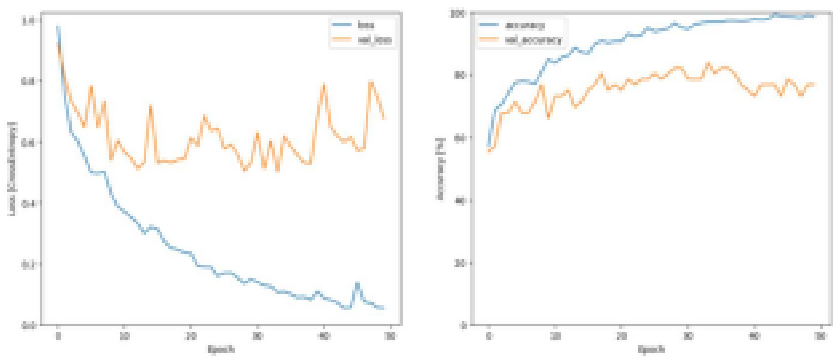

Figure 3. Training Performance Graph

The performance of the multi-class classification model comprising 3 different classes, Happy, Hungry and Scared. The true labels of the test data are shown as the rows, while the predicted labels shown in the columns of the confusion matrix. The values in the cells of the matrix represent the proportion of instances that were correctly or incorrectly classified. The diagonal cells represent the instances that were correctly classified, while the off-diagonal cells represent the instances that were misclassified. The inclusion of the "Unlabelled" class in the confusion matrix reveals that there are 12 instances of unlabelled data. The model correctly predicts all 12 instances as "Unlabelled," demonstrating its effectiveness in identifying instances that do not belong to any specific labelled class. To further evaluate the performance of the model, the researchers calculated three (3) metrics namely individual accuracy, sensitivity, and specificity for each class (see Table 2). These metrics would give a more detailed understanding of how the model is. The sensitivity for the "Unlabelled" class is marked as "Not applicable" since it doesn't have a true positive value, and specificity is not applicable for this class as well. The mean values are calculated by averaging accuracy, sensitivity, and specificity across all classes.

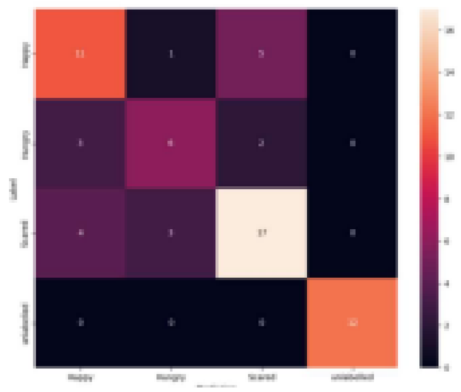

Figure 4. Confusion Matrix of Catsonic

Table 2. Individual accuracy, sensitivity, and specificity performances of each class

Class	Accuracy	Sensitivity	Specificity
HAPPY-	65%	65%	84.6%
HUNGRY	86%	86%	83.8%
SCARED	69%	69%	79.4%
UNLABELLED	100%	NA	NA

The primary focus of this phase was the actual coding tasks for mobile application development. The front-end of the application was developed using the Flutter framework to structure and create user interfaces. Specific functionalities, such as recording and file saving, were implemented using Dart programming language. The entire development process followed established guidelines, with a particular emphasis on creating a user-friendly interface and ensuring the smooth integration of features within the application.

The Permissions and Activation. After pressing the "Listen" button on the Home Page takes users to the Activation and Permissions screen for the microphone. At this crucial point in the user experience, the application makes sure it has access to the microphone on the device so it can record sounds. The system-level permission request shown in the figure asks users to allow access to the microphone.

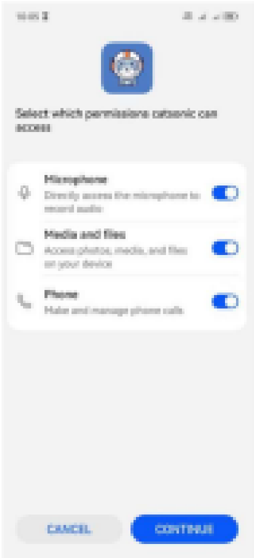

Figure 5. Permissions and Activation of Catsonic

The home page serves as the initial interface of application, offering users a welcoming environment. Here, users encounter a central message displaying "Welcome to Catsonic," signifying a friendly entry to the application. Additionally, a prominent "Listen" button is strategically placed, prompting users to initiate the sound recording process.

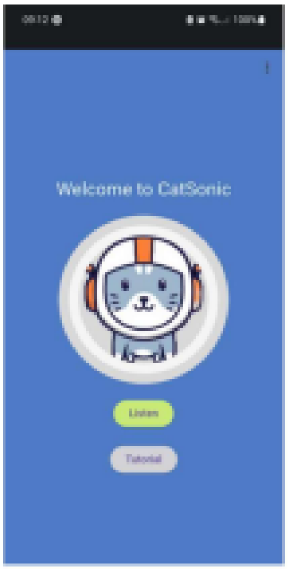

Figure 6. Home Page of Catsonic

Figure 7 presents the launch recording page, after the successful granting of microphone permissions, users are directed to the recording page, a crucial segment for capturing cat sounds effectively. The figure represents the interface during the recording process.

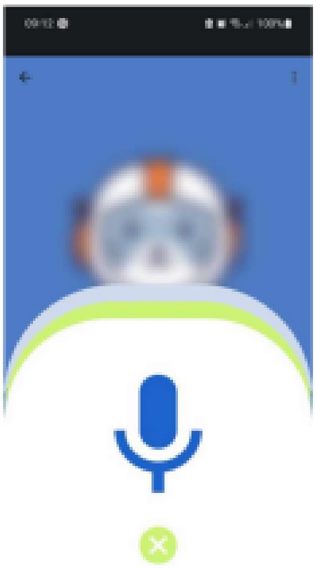

Figure 7. Launch Recording Page

Figure 8 presents the happy classification result. Happy classification result signifies that the recorded sound has been identified as expressing a happy or content mood in a cat. This classification is based on a machine learning model that has been trained on a diverse range of happy cat sounds. Users encounter this page when the application determines a strong likelihood that the recorded sound corresponds to a state of happiness in a cat.

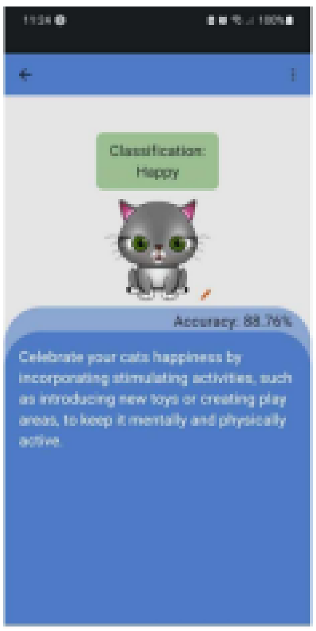

Figure 8. Classification Result

The researchers assessed the usability of the application by providing System Usability Scale through Google Forms to the users. This was done after users have reviewed the application. In the below figure, the statistics were extracted from the Google Form responses to compute the SUS raw and final score. the application garnered a 80.13 out of 100 overall SUS score, which means that the adjective rating is “Excellent”. With this rating, the researchers conclude that the application satisfies the needs of the users.

1	2	3	4	5	6	7	8	9	10		
1. I think that	2. I foun	3. I thought	4. I think	5. I found	6. I thou	7. I wou	8. I fou	9. I felt	10. I nee	SUS Raw Score	SUS Final Score
4	3	4	3	4	3	3	3	3	3	33	82.5
4	4	4	4	4	4	3	4	3	3	37	92.5
3	3	3	3	3	3	3	3	3	3	30	75
4	4	4	4	4	4	4	4	4	4	40	100
4	4	4	4	4	4	4	4	4	4	40	100
3	3	3	2	3	2	3	2	3	2	26	65
4	3	4	4	4	3	4	4	4	3	37	92.5
4	3	3	2	4	4	3	2	4	3	32	80
4	3	4	3	3	2	3	3	4	2	29	72.5
3	2	3	3	3	2	3	3	3	3	25	62.5
3	3	3	3	3	3	3	3	4	3	31	77.5
4	3	4	3	4	3	4	3	4	3	35	87.5
3	3	3	3	3	3	3	3	3	3	30	75
3	2	3	2	3	2	3	2	3	2	25	62.5
2	2	4	2	2	4	3	2	4	2	27	67.5
3	2	3	2	3	2	3	2	3	2	25	62.5
4	3	3	4	4	3	3	4	3	3	34	85
3	3	3	3	3	3	3	3	3	3	30	75
4	4	4	4	4	4	4	4	4	4	40	100
3	3	3	3	3	4	4	4	4	4	35	87.5
											80.126

Figure 9. SUS Computation

5. Conclusion

Based on the objectives and the means the researchers used to accomplish these objectives, the conclusions are as follows: For the first objective, the researchers successfully amassed a diverse set of cat sounds, encompassing both online sources and local recordings. Rigorous labeling procedures were implemented with a consistent naming convention, ensuring systematic categorization. Special attention was given to inclusivity, involving various local cat species to enhance the dataset's richness and representativeness.

For the second objective which is to train the machine learning model for audio classification, the adoption of the transfer learning approach for audio classification proved pivotal in overcoming challenges associated with limited datasets and computational resources. Prudent preprocessing, including percentage splitting and hyperparameter fine-tuning, played a crucial role in optimizing the model's performance and generalization capabilities. This strategic approach laid the foundation for effective machine learning model training.

For the third objective, evaluation of the trained model's performance was conducted through a comprehensive multiclass confusion matrix, offering valuable insights into accuracy, sensitivity, and specificity. The results showed commendable achievements, with an overall accuracy of 80%. However, it’s imperative to acknowledge potential areas for improvement, particularly in addressing instances of misclassification.

Continuous refinement and evaluation are highlighted as essential for enhancing the model's reliability and precision.

For the fourth objective, leveraging the Flutter framework, the researchers successfully developed and integrated the application. This innovative solution recognizes cat vocalizations and classifies. The application goes beyond mere classification, providing users with pertinent information and expert intervention advice tailored to specific vocalizations.

For the fifth objective, the developed application underwent a rigorous evaluation of its usability. The findings revealed a high level of acceptability, with a usability percentage of 80.13%. Users found the application intuitive and user-friendly, affirming its practicality for cat owners and veterinarians alike.

7. References

[1] Naik, P. G. (2023). Introduction to google colab. In Conceptualizing python in google colab: Hands-on practical sessions (p. 1). Shashwat Publication.
[2] Journal of Veterinary Science. (2020). Feline vocal communication. https://doi.org/10.4142/jvs.2020.21.e18.
[3] Ntalampiras, L. A., Ludovico, S., Presti, G., Prato Previde, E. P., Battini, M., Cannas, S., Palestrini, C., & Mattiello, S. (2019). Automatic classification of cat vocalizations emitted in different contexts. Animals, 9 (8), 1. https://doi.org/10.3390/ani9080543
[4] Tavernier, C., Ahmed, S., Houpt, K. A., & Yeon, S. C. (2020). Feline vocal communication. Journal of Veterinary Science, 21(1), 1. https://doi.org/10.4142/jvs.2020.21.e18
[5] ActiveState. (2022, July 11). What is a Keras model and how to use it to make predictions. https://www.activestate.com/resources/quick-reads/what-is-a-kerasmodel/
[6] Bedford, E. (2019, May 7). Number of dogs and cats kept as pets worldwide in 2018. Statista. https://www.statista.com/statistics/1044386/dog-and-cat-petpopulation-worldwide/
[7] Bhandari, P. (2023, June 21). Data collection | Definition, methods & examples. Scribbr. https://www.scribbr.com/methodology/data-collection
[8] Bostech ICT Solutions. (2019, February 20). Impacts of ICT in society. LinkedIn. https://www.linkedin.com/pulse/impacts-ict society-bostech-ict-solutions-bis-/?fbclid=IwAR3_xLl3hAx_hjU0IiL0ZXYWTw9uuwHJiZ5WzMLheIJAkATFch_fJQSxu mA

9. Author(s) Biodata (50 words)

	**SARAH D. OGHENEKARO** is a 4th year undergraduate student at the College of Computer Studies and Engineering of Lorma Colleges. She is interested in data analysis and its practical applications in improving quality of life. Her current research project focuses on sound classification using machine learning. Leisure activities for her include reading and spending time with family. She can be contacted at email: sarah.oghenekaro@lorma.edu.