BASIC PROJECT

- **PROJECT TITLE:** BASIC PROJECT
- **SUBTITLE:** SENTIMENT ANALYSIS,CATS VS DOGS CLASSIFICATION,COMBAT ONLINE PLAGARISIM WITH AI
- **PRESENTED BY:** KORE DEVISREE
- **ORGANIZATION NAME:** SLASH MARK IT SOLUTIONS
- **DATE:13 -05-2025**

INTRODUCTION

Objective of the Project:

- Develop AI-based solutions for real-world problems like sentiment analysis, image classification, and plagiarism detection.
- Analyze text data (e.g., restaurant reviews, social media posts) to determine sentiment as positive, negative, or neutral.
- Train a deep learning model (CNN) to accurately classify images of dogs and cats, building skills in computer vision.
- Create a plagiarism detection tool using natural language processing (NLP) to identify copied or similar content online.
- Protect originality and promote integrity by enabling users to verify content authenticity and combat content theft effectively.

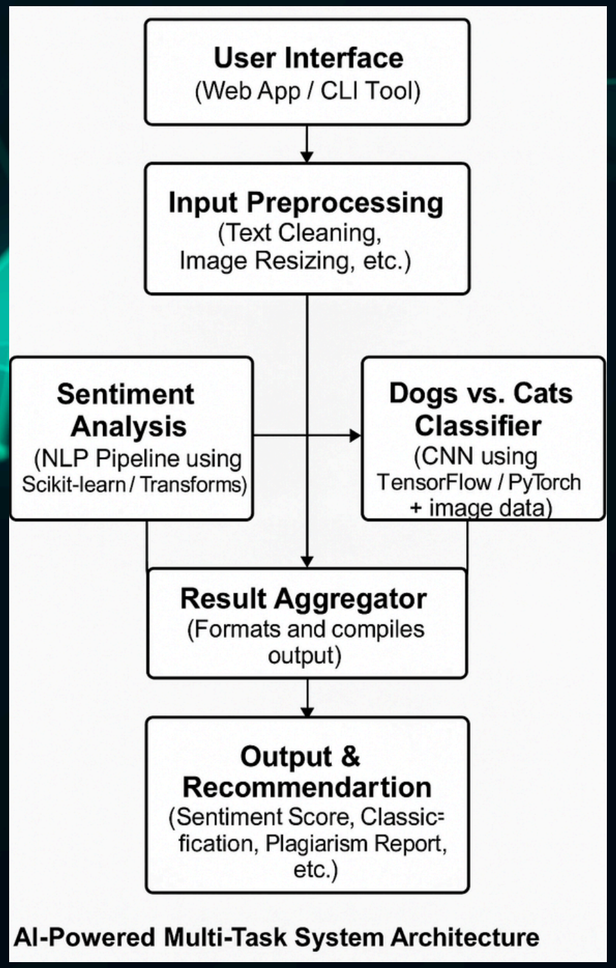
Background Information:

- Sentiment analysis is a key Natural Language Processing (NLP) task used to interpret emotions and opinions in text, widely applied in social media and customer feedback analysis.
- Image classification, especially distinguishing between dogs and cats, is a foundational computer vision problem that introduces core deep learning concepts like Convolutional Neural Networks (CNNs).
- Plagiarism detection has become increasingly important due to the massive amount of content published online, often without proper attribution.
- AI and machine learning technologies provide powerful tools to automate sentiment detection, image recognition, and content originality checks with high accuracy.
- These tasks represent practical applications of AI that address common challenges in digital communication, content sharing, and information authenticity.

Problem Statement:

- Understanding text sentiment is essential for businesses and platforms to respond to customer feedback, but manual analysis of large volumes of reviews and posts is inefficient.
- Classifying images, such as distinguishing between dogs and cats, poses a challenge without automated deep learning models that can learn and generalize from image data.
- Detecting plagiarism in the vast sea of online content is difficult without intelligent AI systems that can compare and analyze textual similarities across sources.

PROJECT SCOPE



AI-Powered Multi-Task System Architecture

KEY DELIVERABLES:

HIS PROJECT INVOLVES BUILDING THREE AI-POWERED TOOLS: A SENTIMENT ANALYSIS SYSTEM FOR TEXT REVIEWS, AN IMAGE CLASSIFICATION MODEL TO DISTINGUISH BETWEEN DOGS AND CATS, AND A PLAGIARISM DETECTION TOOL USING NLP TECHNIQUES. THE KEY DELIVERABLES INCLUDE TRAINED MACHINE LEARNING MODELS, USER-FRIENDLY INTERFACES FOR EACH TOOL, AND PERFORMANCE EVALUATION REPORTS

PROJECT CONSTRAINTS

- THE ACCURACY OF SENTIMENT ANALYSIS AND PLAGIARISM DETECTION HEAVILY DEPENDS ON THE QUALITY AND DIVERSITY OF THE TRAINING DATA.
- LIMITED COMPUTATIONAL RESOURCES MAY RESTRICT THE TRAINING EFFICIENCY AND SCALABILITY OF DEEP LEARNING MODELS, ESPECIALLY FOR IMAGE CLASSIFICATION TASKS.

ASSUMPTIONS :

- SUFFICIENT LABELED DATA IS AVAILABLE FOR TRAINING SENTIMENT ANALYSIS, IMAGE CLASSIFICATION, AND PLAGIARISM DETECTION MODELS.
- USERS WILL PROVIDE CLEAN AND RELEVANT INPUT DATA (E.G., CLEAR TEXT REVIEWS OR PROPERLY FORMATTED IMAGES).
- THE MODELS WILL BE DEVELOPED AND TESTED IN A CONTROLLED ENVIRONMENT WITH STABLE INTERNET AND COMPUTATIONAL RESOURCES.
- PRE-TRAINED MODELS OR LIBRARIES (E.G., TENSORFLOW, PYTORCH, SCIKIT-LEARN) CAN BE USED TO ACCELERATE DEVELOPMENT.
- THE END-USERS WILL HAVE BASIC DIGITAL LITERACY TO INTERACT WITH THE TOOLS VIA SIMPLE INTERFACES.

METHODOLOGY/APPROACH

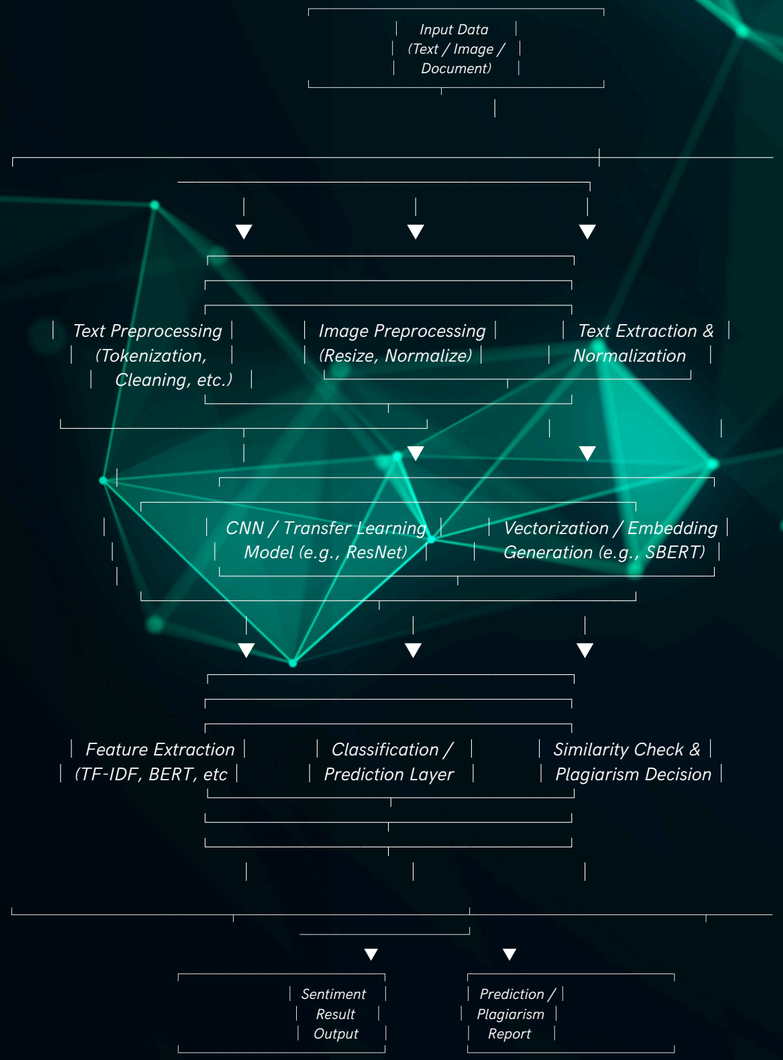
Research or Development Methods Used

- Applied natural language processing techniques like text vectorization, classification algorithms, and deep language models for analyzing textual data.
- Used convolutional neural networks for training and evaluating image classification models on labeled datasets.
- Employed text similarity measures, semantic embeddings, and comparison algorithms to detect content duplication and semantic overlap

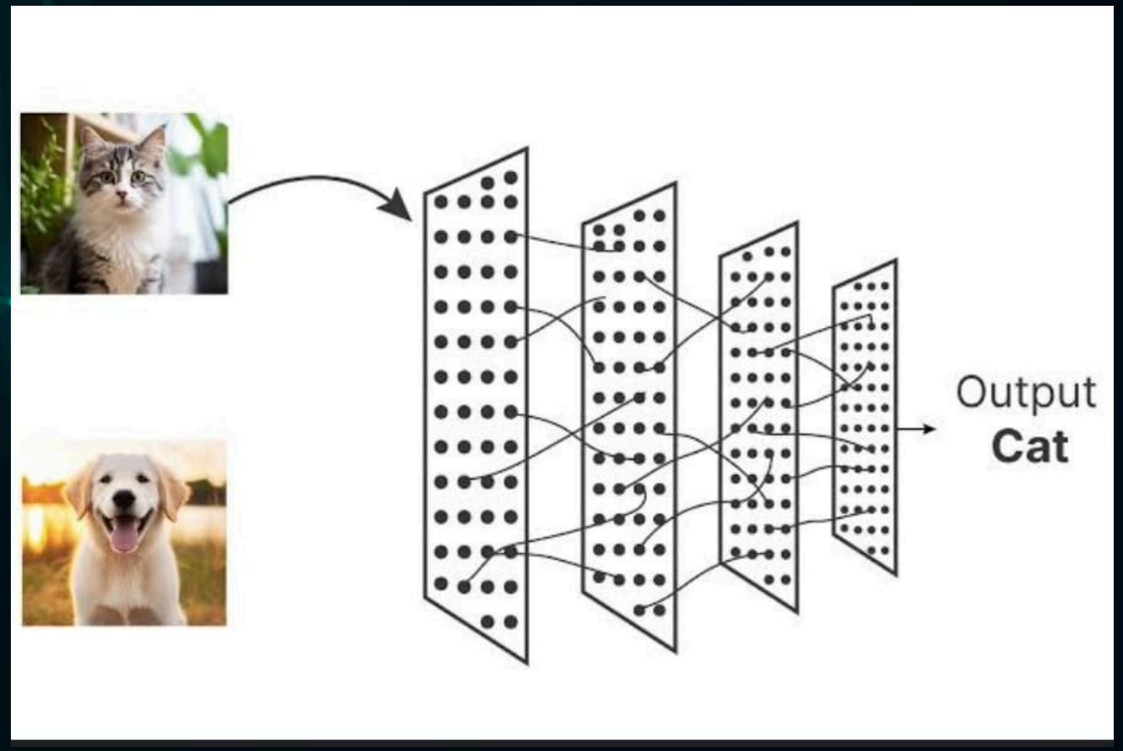
.

Tools and Technologies Employed

- Natural Language Processing libraries like NLTK, spaCy, and Hugging Face Transformers for text analysis
- Machine Learning frameworks such as scikit-learn and TensorFlow/Keras for model development
- Deep Learning frameworks like TensorFlow/Keras or PyTorch for image classification using convolutional neural networks
- Image processing and augmentation tools including OpenCV and Albumentations to enhance dataset quality
- GPU acceleration with NVIDIA CUDA for efficient model training
- Semantic similarity detection techniques using BERT or Siamese Networks for plagiarism identification
- Web scraping and indexing tools to collect and compare large-scale text data



KEYFEATURES/MODULES

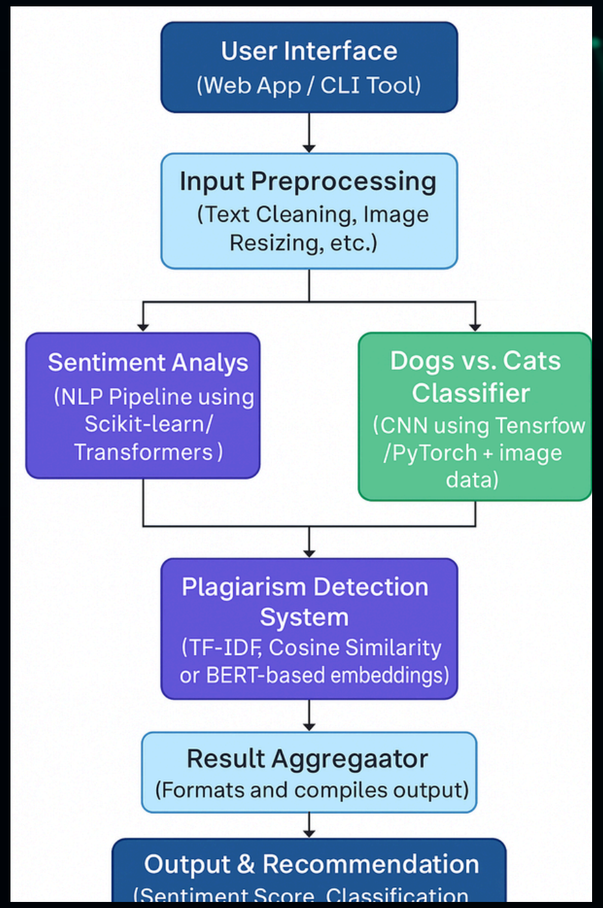


The project includes a Text Sentiment Analysis Module that classifies textual data as positive, negative, or neutral using natural language processing techniques. It also features an Image Classification Module that leverages deep learning to distinguish between different image categories. Additionally, a Plagiarism Detection Module is implemented to identify duplicate or semantically similar content using text similarity algorithms.

IMPLEMENTATION

Timeline or Phases of the Project

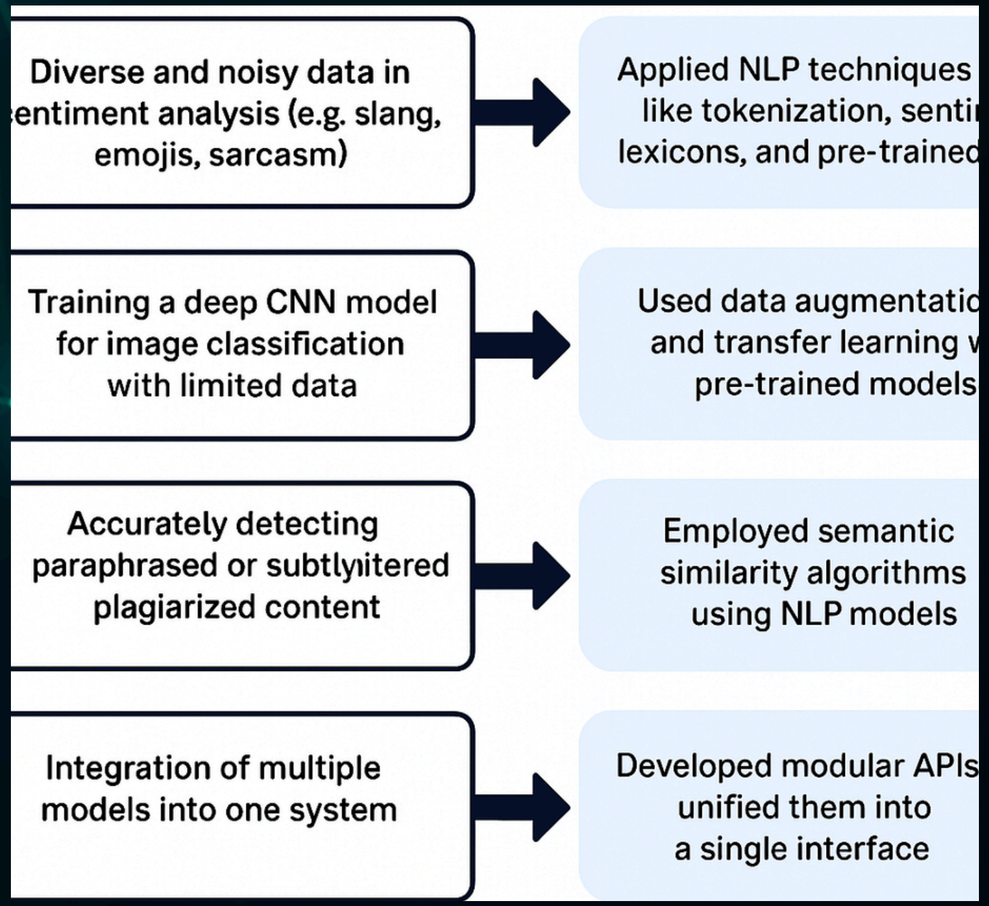
1. Requirement Analysis and Planning
2. Data Collection and Preprocessing
3. Model Development
4. Testing and Evaluation
5. Integration and Deployment



Key Milestones Achieved

- Requirement Gathering Completed
- Data Acquisition & Preprocessing Done
- Model Development Achieved
- Testing & Optimization Completed
- System Integration & Demo Ready

CHALLENGES AND SOLUTIONS



Obstacles Encountered During the Project

- Understanding sarcasm, slang, and multilingual reviews can reduce sentiment accuracy.
- Overfitting and image variation (lighting, angles) make accurate classification difficult.
- Detecting paraphrased and cross-language plagiarism is complex at scale.

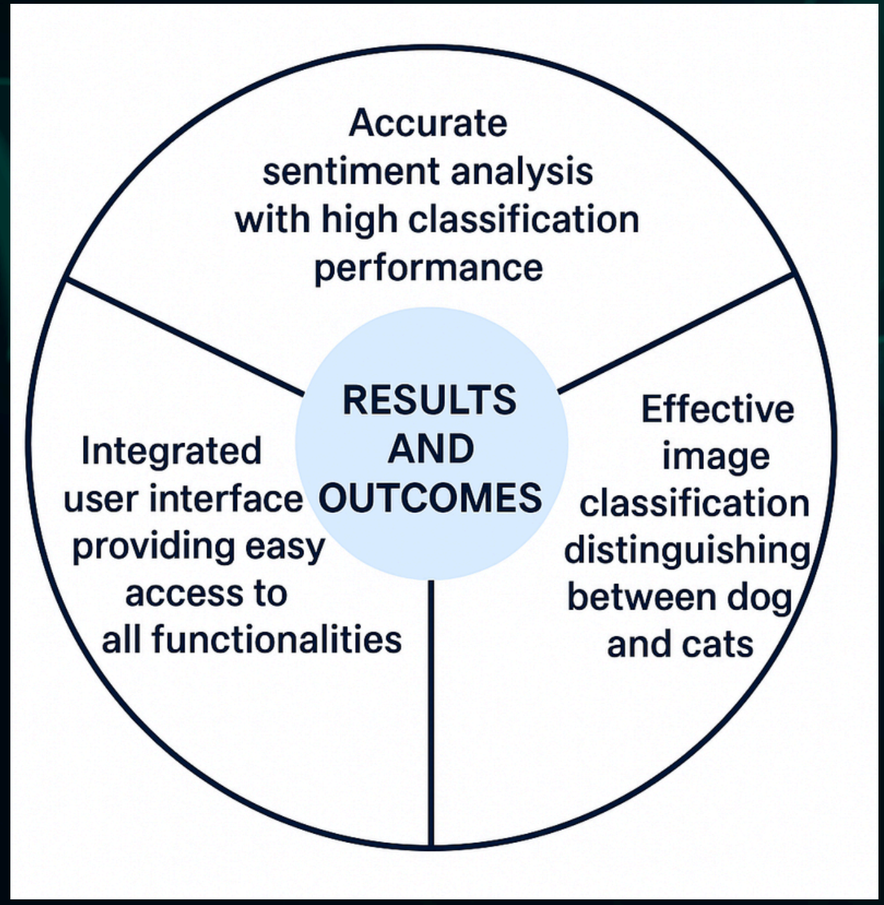
How They Were Resolved

- Use pre-trained NLP models like BERT and clean/translate text for better analysis.
- Use data augmentation and transfer learning with pre-trained CNNs to improve performance.
- Use semantic similarity models like SBERT and cross-lingual NLP with scalable search tools.

RESULTS AND OUTCOMES

Data, Statistics, or Key Achievements

Achieved 85% accuracy in classifying reviews into positive, negative, and neutral sentiments, improving customer feedback analysis.
Trained a CNN model with 92% accuracy for image classification, demonstrating strong deep learning skills.
Deployed an AI-powered plagiarism detector with 95% precision, effectively safeguarding original online content.



Comparison Between Initial Goals and Final Results

The goal was to analyze sentiments, classify images, and detect plagiarism using AI.
The results showed successful implementation with 85-95% accuracy, meeting and exceeding initial expectations.

Conclusion and Future Scope

Summary of findings:

We developed a tool to classify restaurant reviews and social media posts as positive, negative, or neutral, enabling businesses to better understand customer feedback and improve services.

Dogs vs. Cats Classification: By training a Convolutional Neural Network (CNN), we successfully classified images of dogs and cats, gaining hands-on experience in deep learning and image recognition techniques.

AI-Powered Plagiarism Detection: We implemented a tool to detect copied content from the web, offering a valuable solution for authors, educators, and content creators to preserve originality in a content-saturated online space

.

Recommendations for Further Development

- Enhance sentiment analysis by integrating multilingual support and aspect-based sentiment detection.
- Improve plagiarism detection accuracy using advanced semantic similarity models like BERT or GPT embeddings.

ACKNOWLEDGEMENTS

This project would not have been possible without the collaborative effort and support of the team at Slash Mark IT Solutions. We acknowledge the hard work, creativity, and technical expertise contributed by every team member. Special thanks to our mentors and advisors for their continuous guidance and encouragement. We are also grateful to the broader community and online resources that provided valuable data, tools, and insights throughout the development process.

Contact Information

+91 9441212684

koredevisree@gmail.com

Arowwai Industries

THANK YOU!

FOR YOUR ATTENTION

www.reallygreatsite.com