Case Studies in Data Science

WIL Project Group 26 : Written Report

Plant Recognition App

DECLARATION:

"We hereby declare that I have not used generative AI tools to complete this assignment."

Report By:

Kawsar Ali Abdihakim (s3950421)

Mohamed Manzoor (s3914638)

Amirreza Saeidinik (s3932838)

Golnaz Akbari (s3852157)

Contents

Executive Summary 3

Problem Definition 3

Methodology 4

Findings 6

Impact and Significance of Results 8

Conclusion 9

Project Management 9

References 12

Executive summary

The world is on the brink of a botanical catastrophe. A recent survey published in Nature Ecology & Evolution has unveiled a distressing revelation: Human activities spanning centuries have driven a minimum of 571 plant species to extinction in the wild since the 1750s[1]. Compounding this dire situation, alarming data underscores that on average, over eight plant species are vanishing every three years since 1900[1]. This staggering pace of extinction surpasses natural rates by up to 500 times. In this critical moment, we must seize the opportunity to reestablish our connection with the natural world. However, regrettably, many of us struggle to even recognise the plants in our midst, let alone care for them.

Enter our plant recognition app, designed to bridge the gap between humanity and nature. With this tool, we empower individuals to rekindle their bond with the botanical realm, fostering a sense of responsibility towards our planet's irreplaceable plant life.

Our plant recognition app helps in identifying and providing detailed information on various plant species. It provides us with plant names, which type of species, how it should be taken care of and lots of ways to keep the plant healthy and fertilization methods. User only needs to upload a picture of a particular plant through the app. One doesn't require any knowledge about plants to use the app. The app generates accurate results as its model has been trained and fed with different species of plants. Our target audience are:

- Botany students
- Researchers
- Agriculture
- Environmental science
- Forestry
- Gardeners

Problem Definition

Motivation Statement

While many people can readily name an extinct mammal or bird that has become extinct in our last century, few can name a recently extinct plant. The often unnoticed extinction of plants is alarming as they serve as the essential infrastructure of our planet, sustaining life on Earth [1,2]. With the alarming pace of plant extinction continuing to erase plant species from our world, it is now critical to reignite our bond with nature while the opportunity still exists.

Regrettably, many of us find ourselves struggling to identify the plants that surround us, let alone understand how to take care of them effectively. Our plant recognition project addresses this very issue. This project is driven by the desire to reignite our bond with nature. We aim to transform the way people interact with plants and elevate their ability to nurture and care for them effectively.

We are committed to addressing a set of core questions and challenges through this endeavor:

1. Plant Identification:

Our primary challenge is to create a robust system capable of precisely identifying various plant species solely from images. Many individuals struggle with plant identification, whether in gardens, parks, or when encountering unfamiliar plants. Our mission is to provide a user-friendly and dependable solution to this common problem.

2. Personalised Plant Care:

Each plant, whether indoors or outdoors, thrives under specific conditions. Our project seeks to tackle the question of how we can offer users customized guidance on caring for their plants. We understand that sunlight, water, and other factors significantly impact plant health. Our objective is to provide tailored care instructions to ensure optimal growth.

3. Accessibility to Plant Care Knowledge:

Bridging the gap in plant care knowledge is another vital aspect of our mission. We are determined to make plant care information accessible to both seasoned enthusiasts and newcomers. By doing so, we aim to empower individuals with the knowledge they need to cultivate and maintain healthy plants.

4. User Engagement Strategies:

To sustain users' interest and motivation in caring for their plants, we plan to incorporate gamification elements and reminders. Gamification can turn plant care into an engaging and rewarding experience, while timely reminders help users stay on top of their care routines.

5. Continuous Learning:

Beyond offering immediate solutions, our project aims to be an ongoing source of plant knowledge. We intend to provide users with educational content on various plant species and their care requirements. This educational component encourages users to expand their understanding of plants over time.

Methodology

The methodology for addressing the challenges and questions posed in the project primarily involves the utilization of a Convolutional Neural Network (CNN) and associated techniques. Here's a breakdown of the methodology:

1. Curating and Preparing Data:
Our project begins with an extensive dataset encompassing various plant species. Each plant is meticulously categorized by its species. The dataset used was sourced from Kaggle.

Source: https://www.kaggle.com/datasets/muhammadjawad1998/plants-dataset99-classes

To refine and optimize the dataset, a manual curation process was undertaken, this involved eliminating unsuitable plant images and reducing the number of classes. The Kaggle dataset initially consisted of 99 classes, but after our data preparation process it was narrowed down to 52 classes. This was done to improve the accuracy of the model and to reduce the training period required to create the model.

To ensure that the dataset is conducive to training our Convolutional Neural Network (CNN), a series of data preprocessing steps are employed. These steps include tasks like resizing images to uniform dimensions, and applying data augmentation techniques to diversify the dataset.

2. Harnessing the Power of Convolutional Neural Networks (CNN):
In the heart of our methodology lies the selection and implementation of a CNN architecture. CNNs was the chosen method for our project due to its innate ability to excel in image recognition tasks, adept at capturing intricate patterns and nuanced features in images. We embark on the crucial phase of training our CNN model using the meticulously prepared dataset. During this training process, our CNN learns to discern the unique features and patterns that distinguish various plant species from one another.

3. Training Using Transfer Learning with EfficientNetB0:
Transfer learning is often utilized when training a machine learning model to solve a new task that ordinarily would take vast amounts of resources[3]. This process takes relevant parts of an existing machine learning model and applies it to solve a new but similar problem [3]. To simplify, Transfer Learning is when elements of a pre-trained model are reused in a new machine learning model [3]. For our project, we harnessed the power of a pre-trained CNN model known as EfficientNetB0. EfficientNet, initially introduced in Tan and Le, 2019, stands as one of the most efficient models, demonstrating exceptional accuracy on both ImageNet and common image classification transfer learning tasks[4]. This model was selected for our project due to its exceptional performance. EfficientNetB0 consistently outperformed other pre-trained models, ResNet, ImageNet and MobileNet, when tested on this dataset.
Three dense layers were then added on top of the pre-trained model to improve the overall accuracy. Earlier versions of the model had overfitting issues, to reduce bias and improve performance three dropout layers were included.

4. Web Development:
We've designed the app with an intuitive user interface, making it effortless for users to capture plant images using their mobile devices.
The app elegantly presents identification results and care instructions in a manner that is both clear and easy to understand, ensuring that users can seamlessly absorb and apply the provided guidance.

The proposed plant recognition app, leverages Convolutional Neural Networks (CNN) and uses a Django framework to create the web application and deploy it on PythonAnywhere so anyone can access it using the following URL: https://amirrezarmit.pythonanywhere.com/upload/.

The plant recognition app allows users to upload images of plants in a simple way. Once an image is uploaded, it is stored in the database, and the app runs the CNN model for the provided input. It then predicts the name of the plant and displays this information to the user in the web app. Additionally, the app provides care information, offering users guidance on how to best care for the recognized plant. This combination of image recognition and plant care information makes it a valuable tool for plant enthusiasts and gardeners.

Findings

Model Performance

1. Training and Validation Accuracy:

Figure 1 displays model performance through accuracy. Everytime the dataset passes through the algorithm, it is said to have completed an epoch[5]. Accuracy is defined as the percentage of predictions that were made correctly, it is a metric for assessing model performance[6]. In Figure 1 below, note that an accuracy score of 1.00 on the graph represents a 100% accuracy, meaning that for every 10 images shown, all 10 images would be predicted correctly. As displayed in Figure 1, we can see that as the number of epochs increased, so did our accuracy score; however, it is important to highlight that the accuracy began to stabilize around the 35th epoch. In the end, our trained model achieved a notable training accuracy score of 95.79%. This high training accuracy score highlights the models ability for consistent performance.

2. Evaluation Metrics: Precision, Recall, and F1 Score:

Precision denotes the percentage of positive image identifications that were actually correct[6]. Recall denotes the percentage of actual positives that were identified correctly[6]. F1 score is a measure of a test's accuracy, it's a weighted average calculated from precision and recall[6].

In Table 1, we were able to determine the testing accuracy of the model which was evaluated with an accuracy 85.34%. In practical terms, this means that the model accurately identified the correct class of an image in 85.34 out of 100 instances.

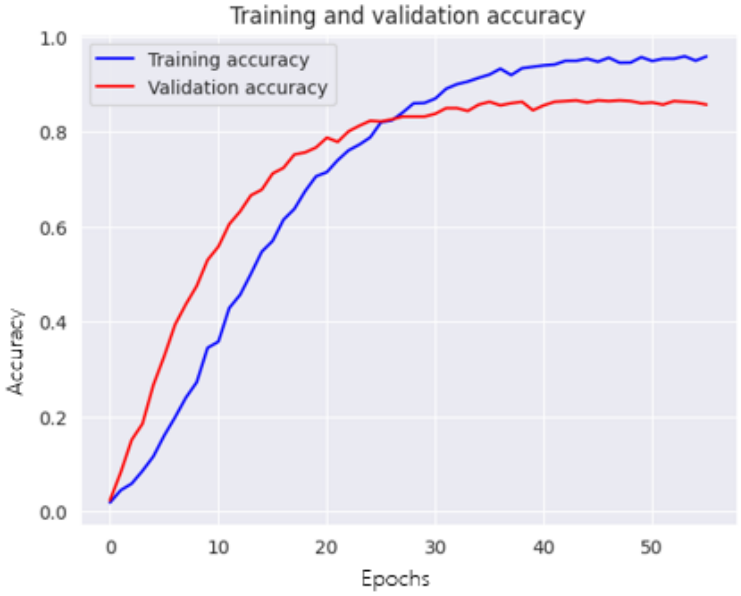

Figure 1. Accuracy Graph depicting Training and Validation Accuracy

	precision	recall	f1-score	support
accuracy	0.853428	0.853428	0.853428	0.853428
macro avg	0.854080	0.835576	0.835955	846.000000
weighted avg	0.858882	0.853428	0.850434	846.000000

Table 1. Classification Report displaying accuracy, recall, precision and F1 score for the model created

Figure 2. Model Loss Graph depicting Training and Validation Loss

3. Training and Validation Loss:

Turning our attention to Figure 2, in the initial stages, both training and validation loss exhibit steep declines during the first 20 epochs;however, after this point, the rate of decrease begins to taper off.

Notably, the validation loss sat slightly higher at the later epochs when compared to the training loss. This gap between validation and training loss suggests some overfitting to the training data. To mitigate this issue, earlier iterations of the model saw the introduction of measures such as reducing the number of classes and incorporating dropout layers. These interventions effectively reduced overfitting concerns, enhancing the model's generalization capability and could be utilized further for a greater reduction.

Given that this model contains 52 classes, it was deemed ineffective to include a confusion matrix in this project. Instead, the report evaluates our project's model by using recall, precision, training

and validation accuracy, training and validation loss, as more suitable metrics to assess model performance.

Looking ahead, it's clear that further refinements are required. Future improvements to the model should aim to optimize its architecture to minimize overfitting. One strategic approach would be to reduce the model's scope, which could involve narrowing the number of considered plant species. By doing so, we anticipate achieving a more robust and generalizable model that can yield improved results in real-world applications.

Impact and Significance of Results

1. Environmental Conservation

The app directly addresses a pressing environmental concern – the rapid extinction of plant species. With over 600 plants having gone extinct in the past 250 years, the app's ability to identify, care for, and conserve plants is of paramount importance. It contributes to the preservation of plant diversity, thereby safeguarding ecosystems.

2. Educational Empowerment

The app offers significant educational value to users, catering to botany enthusiasts, gardeners, and anyone curious about the plant world. It fosters a deeper understanding of plant life, nurturing an appreciation for the natural environment. This knowledge can drive sustainable practices and conservation efforts.

3. Enhanced Efficiency for Professionals

Plant ecologists and researchers stand to benefit greatly from the app. It streamlines their work by enabling quick and accurate plant identification, which enhances data collection and monitoring efforts. Such efficiency is essential for informed decision-making in environmental management and conservation.

4. Enriching Travel Experiences

The app's traveler-oriented feature adds value to wanderers by allowing them to identify and learn about the plants they encounter during their journeys. This adds an educational and insightful dimension to travel, enhancing the quality of their experiences and fostering a deeper connection to nature.

5. Monetization Opportunities

From a business perspective, the app presents various monetization avenues, including premium features, advertising, and partnerships with botanical organizations and nurseries. These revenue streams not only sustain the business but also expand the range of services available to users.

6. Promoting Sustainability

By encouraging responsible plant care and increasing awareness about diverse plant species, the app contributes to a more sustainable and environmentally conscious society. It aligns with the global movement towards conservation and sustainability, helping build a greener future.

7. Community Building

The app can facilitate the formation of a vibrant community of plant enthusiasts, gardeners, and nature lovers. This community can exchange knowledge, share experiences, and offer valuable tips, further enhancing the app's value and strengthening the mission of environmental conservation.

8. Long-Term Vision

The app's aspiration to strengthen the human-plant connection aligns perfectly with the global trend towards environmental stewardship and sustainability. Its long-term vision can create lasting value by contributing to a greener, more harmonious future, fostering a deeper appreciation for the natural world.

Conclusion

In a world where our connection to the natural environment is dwindling, our plant recognition app, powered by a CNN model and built on the Django framework, emerges as a vital bridge for those who yearn for a deeper connection with nature. The model has succeeded in achieving remarkable results, with a training and test accuracy of 95.79% and 85.34% respectively, this app excels in plant identification. With the help of our plant recognition app not only do we provide users with the names of plants but we also impart essential information on plant care. In essence, our plant recognition app is not merely a technological advancement; it is a chance for us to reconnect with nature. As we continue to enhance its capabilities with improvements to the model and additional features, our project holds the potential to evolve into an even more refined and impactful tool for our planet's future.

Project Management

Tools utilised for communication purposes include Microsoft Teams, Google Meet and WhatsApp. For planning, delegations of tasks, and timekeeping, we utilised a Gantt chart that was hosted on Google Sheets. For collaborating on our written reports and initial brainstorming we utilised Google Docs. Please see below Figure 3 for the Gantt chart used for this project, also see Table 2 displaying the breakdowns used.

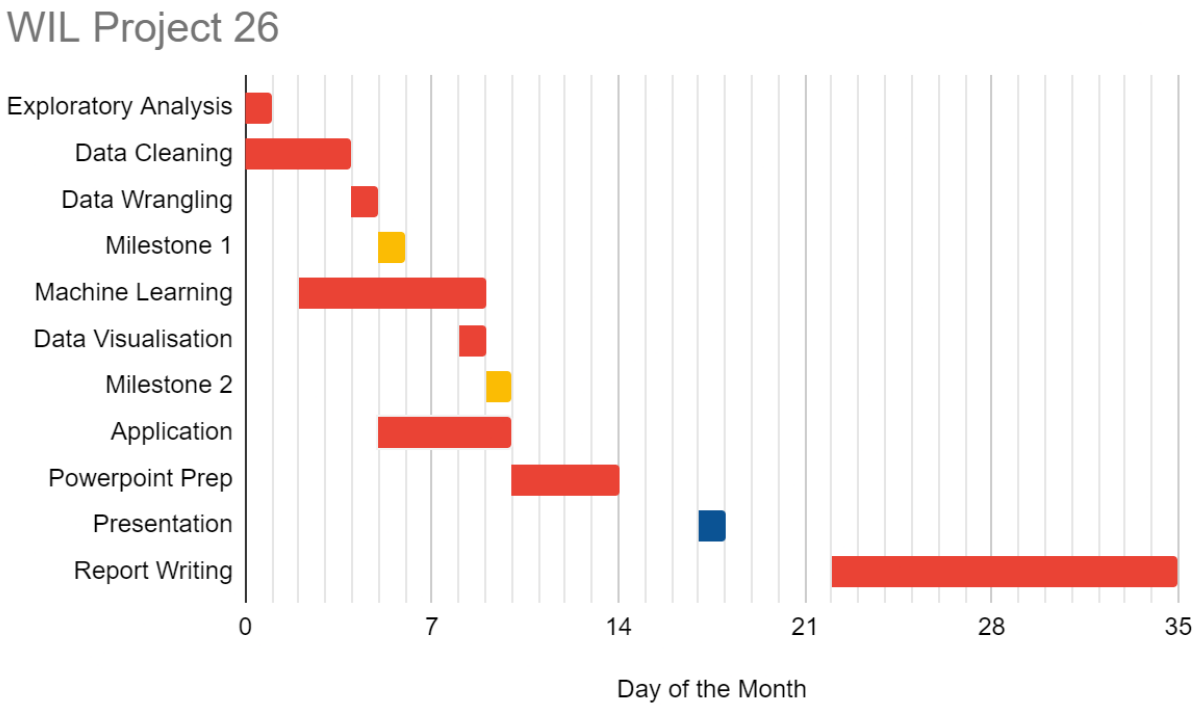

Figure 3. Gantt Chart used for management of this project, the chart visualises duration and project timeline over a period of time

Table 2. Table was used to generate our Gantt Chart, the table displaying roles, duration and our timeline over a period of time

Assignee	Supervisor	Task Name	Expected Starting Date	Expected Ending Date	Start on Day	Duration	Status
Kawsar	Amir	Exploratory Analysis	15/09/2023	16/09/2023	0.00	24:00:00	Completed
Amir	Golnaz	Data Cleaning	15/09/2023	19/09/2023	0.00	96:00:00	Completed
Golnaz	Manzoor	Data Wrangling	19/09/2023	20/09/2023	4.00	24:00:00	Completed
All		Milestone 1	20/09/2023	21/09/2023	5.00	24:00:00	Completed
Kawsar & Golnaz	Amir & Manzoor	Machine Learning	17/09/2023	24/09/2023	2.00	168:00:00	Completed
Manzoor	Kawsar	Data Visualisation	23/09/2023	24/09/2023	8.00	24:00:00	Completed
All		Milestone 2	24/09/2023	25/09/2023	9.00	24:00:00	Completed
Amir & Manzoor	Kawsar & Golnaz	Application	20/09/2023	25/09/2023	5.00	120:00:00	Completed

All		Powerpoint Prep	25/09/2023	29/09/2023	10.00	96:00:00	Completed
All		Presentation	02/10/2023	03/10/2023	17.00	24:00:00	Completed
All		Report Writing	07/10/2023	20/10/2023	22.00	312:00:00	Completed

References

[1] J. P. Wright, S. Naeem, A. Hector, G. Bell, D. U. Hooper, M. Loreau, P. B. Reich, M. Solan, and D. Tilman, "Biodiversity loss and the functioning of ecosystems: Evidence from experimental data," Nature, vol. 466, no. 7302, pp. 1093–1098, 2019.

[2] National Geographic, "Plant extinctions are shaking Earth's green foundation," National Geographic,2023.[Online].
Available:https://www.nationalgeographic.com/environment/article/plant-extinctions-are-shaking-earths-green-foundation#:~:text=But%20the%20biodiversity%20crisis%20isn,%2C%20on%20average%2C%20since%201900. [Accessed: 10 October 2023].

[3]Seldon,"Transferlearning," Seldon, 2023. [Online]. Available:https://www.seldon.io/transfer-learning#:~:text=Transfer%20learning%20is%20generally%20used,categorisation%20or%20natural%20language%20processing. [Accessed: 10 October 2023].

[4] Keras, "Image Classification with EfficientNet Fine-Tuning on TF-Hub," TensorFlow, 2023. [Online].Available:https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/.[Accessed:11 October 2023].

[5] Simplilearn, "What is Epoch in Machine Learning?" Simplilearn, 2023. [Online]. Available: https://www.simplilearn.com/tutorials/machine-learning-tutorial/what-is-epoch-in-machine-learning. [Accessed: 11 October 2023].

[6] S. Raschka, "Choosing Performance Metrics," Towards Data Science, 2020. [Online]. Available: https://towardsdatascience.com/choosing-performance-metrics-61b40819eae1. [Accessed: 12 October 2023].