Team Leader – Milestone 1

Kenneth Mcferrier (577503)

Team Members:

Leon Botha (576378)

Keneilwe Mpho Ramatsui (577669)

Kaylee Carstens (577331)

Kealeboga Molefe (577482)

Keletso Masoho (577546)

PRJ381: Plant Phenology Identification

GroupID: PRJ381-35

Table of Contents

Executive Summary

This project uses visual data to identify and categorise the phenological stages (bud, flower, and fruit) of Aloe Ferox plants using machine learning techniques. In the end, this project will yield a user interface for data visualisation and interaction, a machine learning model with accurate phenological categorization capabilities, and tools for exporting and analysing prediction data. The project aims to shed light on the phenology of Aloe Ferox and help comprehend its growth patterns and responses to climatic circumstances by examining these stages and linking them to environmental factors. The main deliverables comprise of the trained model, a working user interface, and thorough documentation of the results and techniques utilised.

Project Objectives

- **Identification**: Determine whether buds, flowers, or fruit (or any combination of these) are present in a photo of an Aloe Ferox plant.
- **Tagging**: Assign the appropriate phenology classification label to each image in the spreadsheet/database.

- **Environmental Analysis:** To determine the meteorological and geographic factors impacting the phenology stages, use date, seasonal, and location data.

Project Scope

Included:

Image collection:

- Collect and prepare Aloe Ferox photos from resources such as iNaturalist, making sure the labels are appropriate for training.

Dataset Creation:

- Using labelled data from the spreadsheet and associated images to create a training dataset for a machine learning algorithm.

Algorithm Selection:

- Selecting a suitable machine learning method to recognize and classify Aloe Ferox's distinct flowering stages (bud, flower, and fruit).

Model Validation:

- To make sure the trained model is accurate and reliable, test it on a different dataset.

Analysis of Environmental Data:

- To examine the connection between environmental influences and phenological stages, gather weather and geographic information from resources like Google Earth Engine and the South African Weather Service.

Development of the User Interface:

- Provide a user-friendly interface that makes it easy to upload photos, view categorization results, and visualise phenological data.

Data Exporting Capability:

- Add functionality that allows you to export the model's accuracy estimates and forecasts to an Excel file for additional study.

Excluded:

Utilisation in Other Plant Species:

- Adapting the model without more data gathering or model training to classify other plant species' phenological stages.

Extended Observation:

- Monitoring the phenology of Aloe Ferox even after the project is finished.

Requirements Analysis

Functional Requirements:

Image Collection and Labelling:

- Collect images of Aloe Ferox from iNaturalist.
- Label images based on phenological stages (bud, flower, fruit).

Data Management:

- Import labeled data from a spreadsheet.
- Export processed data to an Excel file with accuracy estimates.

Machine Learning Model:

- Select and implement a suitable machine learning algorithm.
- Train the model using labelled data.
- Validate the model's accuracy on a test dataset.

Prediction and Analysis:

- Use the model to predict phenological stages from new images.
- Analyse date, seasonal, and location data to determine environmental impacts on phenology.

User Interface:

- Develop a user-friendly interface for data input and display.
- Enable users to load images and view phenological classifications.
- Provide tools for exporting prediction data and accuracy estimates.

Non-Functional Requirements:

Performance:

- Ensure the machine learning model provides accurate and timely predictions.
- Optimize data processing and model training times.

Usability:

- Design an intuitive and easy-to-use interface for users with varying technical backgrounds.

Scalability:

- Ensure the system can handle a large number of images and associated data without performance degradation.

Security:

- Protect data integrity and privacy, especially when dealing with location-specific information.

Maintainability:

- Write clean, well-documented code to facilitate future updates and maintenance.
- Provide clear guidelines for adding new data and retraining the model.

High Level design

Major Components:

Hardware Components:

Server Infrastructure:

- Used for processing, storing, and training models.
- Offers the computing capacity needed to manage big datasets and complex math.

Adequate Computers:

- They will be used for training and running the machine learning model.

Networking Equipment:

- The routers and switches that ensure proper network communication between servers and with external data sources.

Software Components:

Data Collection and Storage:

- Data Collection Scripts for fetching images and metadata from iNaturalist.
- Data Collection Scripts for collecting weather data.
- Database management system (DBMS) for storing data.

Data Processing:

- Software for preprocessing images, such as augmenting, normalisation, and scaling.
- Tools for labelling data using the supplied spreadsheet as a guide.

Machine Learning Component:

- Machine learning software and tools like TensorFlow for training and developing models.
- Algorithms such as Convolutional Neural Networks for classifying images.
- Module for generating phenological stage predictions.

User Interface:

- For users to upload images for classification.
- Displays predictions and associated data.
- Allows exporting prediction results to Excel.
- Shows interactive data visualizations.

Exporting:

- The accuracy estimates and classification results can be exported by users to an Excel file for additional examination.

Deployment and Maintenance:

- Scripts for deploying software components to servers or cloud.
- Tools for system updates, model retraining, and operational maintenance.

Interactions Between Software Components:

- **Data Collection and Storage to Data Processing:** The data gets collected and stored in a DBMS where it then gets processed and labelled as needed.
- **Data Processing to Machine Learning:** The processed and labelled data will then be used by the Machine Learning model to train and validate it.

- **Machine Learning to User Interface:** Through the user interface, users upload new images, which the trained model then classifies. Visuals of the study of the environmental data are presented alongside the results.
- **User Interface to Exporting:** The accuracy estimates and classification results can be exported to an Excel file for additional examination.
- **Exporting to Deployment and Maintenance:** After exporting the results the maintenance phase can begin to ensure that model retraining can be done if needed and to push system updates.

Overall Design Approach:

Modular Design:

Because each part is made to be a standalone module, independent testing, development, and maintenance can take place. Modular design also makes updating and the integration of new features a lot easier.

Interoperability:

By using standard APIs, you can be assured that components can connect with each other and with additional data sources or analytics tools down the road.

User Friendly Design:

The web interface offers simple navigation and dynamic visuals to improve user experience. It is designed to be intuitive and user-friendly.

Scalability:

By scaling up resources like database capacity and analytics computing power, the architecture allows the system to adapt to growing data volumes and processing needs.

Security:

To protect private information and preserve sensitive data, the system has access limits, secure APIs, and data encryption built in.

Project Timeline

Milestone 1 – Planning 2024-05-20 to 2024-07-20

Week 1-2 (2024-05-20 to 2024-06-02)

- Conduct project kick-off webinar (2024-05-13).
- Define detailed project scope and objectives.
- Identify and finalize stakeholders' roles and responsibilities

Week 3-6 (2024-06-03 to 2024-06-30)

- Collect Aloe Ferox images from iNaturalist.
- Begin data labelling process for phenological stages (bud, flower, fruit).
- Acquire necessary environmental data sources (weather, location).

Week 7-8 (2024-07-01 to 2024-07-14):

- Finalize dataset preparation and initial data analysis.
- Select suitable machine learning algorithms for phenology classification.

Week 9 (2024-07-15 to 2024-07-20):

- Complete planning documentation.
- Prepare for Milestone 1 review and approval by stakeholders.

Milestone 1 Release 2024-07-21

Milestone 2 - Design 2024-07-22 to 2024-09-05

Week 10-11 (2024-07-22 to 2024-08-05):

- Design machine learning model architecture.
- Implement initial image processing algorithms.

Week 12-14 (2024-08-06 to 2024-08-27):

- Preprocess training dataset and integrate labelled data.
- Develop user interface prototype for image data handling.

Week 15-16 (2024-08-28 to 2024-09-05):

- Conduct model training and optimization.
- Finalize user interface design and integration with backend systems.
- Prepare documentation for Milestone 2 review and approval.

Milestone 2 Release 2024-08-06

Milestone 3- Implementation, Test & Demo 2024-09-16 to 2024-10-31

Week 17-19 (2024-09-16 to 2024-10-06):

- Implement machine learning model into production environment.
- Conduct rigorous testing and validation of the model.

Week 20-21 (2024-10-07 to 2024-10-21):

- Develop demo scenarios and prepare presentation materials.
- Finalize project documentation and user manuals.

Week 22-23 (2024-10-22 to 2024-10-31):

- Conduct final project demo and presentation to stakeholders.
- Gather feedback and finalize any outstanding tasks.
- Prepare for project closure and handover.

Milestone 3 release 2024-11-01

Week 24 (2024-11-04 to 2024-11-08)

- Expo Preparation

Methodology

Agile Methodology

The project will be implemented using agile principles to ensure flexibility and adaptability during the execution phase. Agile development will encourage continuous improvement of processes and results, iterative development, and continuing feedback.

Processes

- **Compiling all the requirements:** To determine the requirements, there has to be talks with the stakeholders, supervisor and with team members.
- **Development Phases:** Conduct regular testing and validation while implementing functionality.

- **Analysis and Recommendations:** Continuous communication between other team members and supervisor. Taking feedback into consideration so the best possible outcome can be reached.

Stakeholder involvement and roles

- **Project supervisor:** Throughout the project's duration, the project supervisor should provide oversight, guidance, and constructive criticism.
- **Project sponsor:** This person is responsible for ensuring the project's success and providing the team and project manager with enough resources and direction.
- **Technical support:** They provide guidance on machine learning and GIS tools, as well as assistance with technical issues.

Relevant Documentation

- **Product backlog:** An exhaustive inventory of every feature, enhancement, and bug fix required for the project.
- **Sprint backlog:** The portion of the backlog devoted to tasks to be completed in a particular sprint.
- **Daily stand-up reports:** Short summaries of the team's goals, difficulties, and advancement.
- **Sprint planning:** Outlines the goals and assignments for every sprint.

Relevant Templates and/or forms

Product Backlog with relevant features:

Product Backlog with relevant features

Task	Priority	Category
Collect Aloe Forex Images from iNaturalist and other resources.	High	Data Collection
Obtain location, date, and weather data for each image.	High	Data Collection
Ensure data quality and consistency.	High	Data Collection
Standardize image sizes, enhance features, and remove noise.	High	Data Preprocessing
Extract relevant features from images.	High	Data Preprocessing
Categorize images based on phenological stages (bud, flower, fruit).	High	Dataset Preparation
Divide the dataset into training and testing sets.	High	Dataset Preparation
Generate synthetic data if necessary to balance the dataset.	High	Dataset Preparation
Choose suitable machine learning algorithms.	High	Model Selection
Train models on the labeled dataset.	High	Model Training
Optimize model parameters for best performance.	High	Model Training
Assess model accuracy, precision, recall, and F1 score.	High	Model Evaluation
Design a user-friendly interface for image upload and display.	High	User Interface Development
Integrate the trained model with the user interface.	High	User Interface Development
Develop functionality to display phenological predictions.	High	User Interface Development
Enable export of prediction results to Excel files.	High	User Interface Development
Integrate weather data for each image location.	Medium	Data Analytics
Analyze phenological stages in relation to weather patterns.	Medium	Data Analytics
Create visualizations to illustrate budding, flowering, and fruiting stages.	Medium	Visualization
Generate reports summarizing model performance and phenology insights.	Medium	Reporting
Validate the model on a separate holdout dataset.	Medium	Model Validation
Perform cross-validation to ensure model robustness.	Medium	Model Validation
Analyze prediction errors to identify improvement areas.	Medium	Model Validation
Further optimize and fine-tune the model.	Low	Optimization
Enhance UI/UX based on user feedback.	Low	UI/UX Enhancement
Deploy the machine learning model and user interface.	Low	Deployment and Maintenance
Implement monitoring to track model performance in production.	Low	Deployment and Maintenance
Plan for ongoing updates and improvements to the model and UI.	Low	Deployment and Maintenance
Define sprint goals and tasks.	Low	Project Management
Conduct regular team check-ins.	Low	Project Management
Demonstrate progress to stakeholders.	Low	Project Management
Reflect on sprint performance and identify improvements.	Low	Project Management
Maintain comprehensive project documentation.	Low	Documentation

Sprint Planning Template:

Sprint Planning

Sprint number	Topic	Start Date	End Date
1	Project Kickoff and Scope Definition	May 20	Jun 2
2	Data Collection and Initial Labeling	Jun 3	Jun 16
3	Continued Data Labeling and Acquisition	Jun 17	Jun 30
4	Dataset Preparation and Algorithm Selection	Jul 1	Jul 14
5	Planning Documentation and Review Preparation	Jul 15	Jul 20
6	Model Architecture and Initial Processing	Jul 1	Jul 14
7	Data Preprocessing and Integration	Jul 15	Jul 28
8	User Interface Prototype Development	Jul 29	Aug 11
9	Model Training and Optimization	Aug 12	Aug 25
10	Finalizing UI and Integration	Aug 26	Sep 6
11	Model Deployment	Sep 16	Sep 29
12	Rigorous Testing and Validation	Sep 30	Oct 6
13	Demo Preparation and Documentation	Oct 7	Oct 20
14	Final Demo and Project Closure	Oct 21	Oct 31
15	Expo Preparation	Nov 4	Nov 8

Daily Standup Meeting Template:

Daily Standup Meeting

Team member	Date	Tasks done Yesterday	Tasks to do Today	Challenges
Kenneth	Jul 22	None	Start with collecting Images and data.	Struggled to log into iNaturalist for a while.
Leon	Jul 22	None	Label the initial data.	None

Requirements for daily standup meetings:

- **Participants:** This includes the scrum master and development team.
- **Consistent format:** There are three key questions that need to be answered every time which consist of the following:
  1. What did I do yesterday?
  2. What am I going to do today?
  3. What obstacles will stand in my way?

- **Time schedule:** Daily standup meetings will take place at 9 in the morning every Monday to Friday and will last for 15 minutes give or take a few minutes.
- **Objectives:** All of the team members need to know and understand what the objectives of the whole project and different relevant sprints are.

Resource Management

Personnel

Project Team:

- **Project Manager:** Responsible for overseeing the project's coordination, scheduling, and stakeholder communication.
- **Scrum Master:** Manages challenges, oversees scrum procedures, and guarantees adherence to agile principles.
- **Data Analyst:** In charge of gathering images.
- **Machine Learning Engineer:** In charge of developing the Machine Learning Model.
- **Software tester:** Examines the system to find any flaws or deficiencies.
- **User Interface Designer:** This person creates the user interface and makes sure it is extremely user-friendly.

Equipment

Hardware:

- Workstation: High Performance Computers and/or Laptops for each group member
- A on-premises server: Dual Intel Xeon
- Cloud server: Google Cloud
- Peripheral devices: A monitor, external storage and network equipment

Software:

- Data Collection Tool: iNaturalist
- Machine Learning Tool: TensorFlow
- User Interface Tool: Tableau
- Exporting Tool: Excel
- Testing Tool: Junit
- Project Management Tool: Monday

Materials

- Images from iNaturalist.

Budget

Total cost breakdown:

Personnel costs:

The average salary in South Africa for each of the Team members varies and might look like the following:

(Note that this is just a rough estimate)

- Project Manager (Talent.com, 2024): R308/hour x 20 hours/week x 20 weeks = R123200
- Scrum Master (Talent.com, 2024): R277/hour x 20 hours/week x 20 weeks = R110800
- Data Analyst (Talent.com, 2024): R246/hour x 20 hours/week x 20 weeks = R98400
- Machine Learning Engineer (Talent.com, 2024): R418/hour x 20 hours/week x 20 weeks = R167200
- Software Tester (Talent.com, 2024): R200/hour x 20 hours/week x 20 weeks = R80000
- User Interface Designer (Talent.com, 2024): R267/hour x 20 hours/week x 20 weeks = R106800

Total Personnel costs = R686 400

Equipment costs:

- High Performance computers and/or Laptops (Evetech, 2024): R15000 each x 6 = R90000
- TensorFlow: Free
- Tableau (Tableau, 2024): R1 370
- Dual Intel Xeon: Free
- Microsoft Excel: Free

- Junit: Free
- Monday: Free
- Peripheral devices: Free (Already own them)
- iNaturalist: Free

Total Equipment costs = R91 370

Total Costs = R777 770

Emergency Fund:

15% of Total costs = R116 665,5

Total Budget:

Total cost + Emergency fund = R894 435,5

Risk Management

Potential Risks

- There could be an insufficient or low-quality data from iNaturalist or other sources
- The selected machine learning algorithm may not accurately identify the phenological stages of Aloe Forex
- The model may overfit or underfit
- Inaccurate or incomplete weather data could lead to incorrect correlations between phenology and environmental factors.
- The user interface may not meet the usability expectations of stakeholders
- There could be delays in data collection, model development and unforeseen technical challenges

Mitigation Strategies

- Conduct thorough data quality checks before using the data
- Implement cross-validation techniques to assess model performance
- Ensure sufficient training data and adjust model complexity accordingly
- Validate weather data from reliable sources
- Involve stakeholders in the UI design process for feedback and iterative improvements.

- Develop a detailed project plan with milestones and deadlines.
- Regularly monitor progress and adjust timelines as needed

References

Evetech. (2024, July 20). *Lenovo IdeaPad Gaming 3 15ACH6 82K200U2SA AMD Ryzen 5 5600H up to 4.20GHz Processor, 16MB Cache, 6x Cores, 12x Threads / 8GB DDR4 RAM / 256GB Ultra-Fast NVMe SSD + 1TB Hard Drive / 15.6" FHD 1920 x 1080 Anti-Glare 120Hz IPS-Level Display / NVIDIA GeForce* . Retrieved from Evetech: https://www.evetech.co.za/lenovo-ideapad-gaming-3-15ach6-ryzen-5-rtx-3050-laptop/laptops-for-sale/29098.aspx

Tableau. (2024). *Pricing for data people*. Retrieved from Tableau: https://www.tableau.com/pricing/individual

Talent.com. (2024). *Data Analyst average salary in South Africa, 2024*. Retrieved from Talent.com: https://za.talent.com/salary?job=data+analyst

Talent.com. (2024). *Machine Learning Engineer average salary in South Africa, 2024*. Retrieved from Talent.com: https://za.talent.com/salary?job=machine+learning+engineer

Talent.com. (2024). *Project Manager average salary in South Africa, 2024*. Retrieved from Talent.com: https://za.talent.com/salary?job=project+manager

Talent.com. (2024). *Scrum Master average salary in South Africa, 2024*. Retrieved from Talent.com: https://za.talent.com/salary?job=scrum+master

Talent.com. (2024). *Software Tester average salary in South Africa, 2024*. Retrieved from Talent.com: https://za.talent.com/salary?job=software+tester

Talent.com. (2024). *Ui Designer average salary in South Africa, 2024*. Retrieved from Talent.com: https://za.talent.com/salary?job=ui+designer