

Introduction

Computerised devices are being used in every part of our lives. Our digital assets are vulnerable to attack from malicious files. There are several file types and extensions that can be dangerous to our computer system (Hoffman, 2017) which may be difficult for an average computer user to differentiate between ordinary and malicious files. The aim of our project is to deploy intelligent agent to automate the process of classifying files. This project design proposal provides an outline plan for designing and developing an intelligent agent to classify files based on the file name extension.

Objectives

The main objectives of our project are:

* To design and develop an intelligent agent to be able to analyse the file name extension for classification with various file types such as documents, images, audio and video files (Examples: JPEG, PNG, PDF, DOCX, XLXS, MP3, WAV, etc.
* To implement machine learning algorithm to train agent to recognise the file names accurately
* To develop an interface for users to interact with the agent to review and validate the classification results and provide feedback.
* To perform testing and validation to ensure accuracy and efficiency of the intelligent agent

Methodology

Requirement Analysis

Defining file types: Define the file types and extensions to be classified

Data preparation: Identifying features ( file name extension)

Choice of Python libraries:

Choose Machine learning algorithm for classification: decision trees?

Project Design and Architecture (Approach and Rationale)

using the CRISP-DM methodology provides a structured, adaptable, and best-practice-driven approach to data mining projects. It helps ensure project success, effective collaboration, and the delivery of valuable insights and outcomes for organisations. (Ref)

Two different graphical designs – UML model

Development and Implementation

Split into training and testing data (technique)

Which development techniques and areas are you leveraging

Testing and Evaluation

Cross-validation technique? Overfitting?

Evaluation: Use evaluation metrics: accuracy, precision, recall, and F1-score. Analyze the results to identify any misclassifications or areas for improvement. Adjust the model or experiment with different techniques, such as feature engineering or ensemble methods, to enhance its performance.

Deployment: interface or API that allows users to input files, and the intelligent agent will provide predictions or classifications based on the trained model.

Timeline

Requirement Analysis: 5 days

Design and Architecture: 10 days

Development and Implementation: 15 days

Testing and Evaluation: 10 days

Documentation: 5 days

Resources

3-member student team

Python libraries

Datasets for training and testing

Different ML models can be integrated into agent based model to improve results (Brearcliffe 2021)

Challenges encountered: various challenges you have identified/expect to encounter when developing your agent. E.g., are there any issues around identifying and retrieving the data? Highlight (briefly) what paradigm(s), pattern(s), theories and practices you intend to utilise on this project to address the challenges. Justify your approaches supported by literature.

Expected Outcomes

A functional intelligent agent capable of classifying files accurately based on their file name extension.

Improved data organisation and retrieval efficiency for users, reducing manual efforts in file management.

Conclusion

This project proposal outlines the plan for designing and developing an intelligent agent focused on file classification. By implementing machine learning techniques, the agent aims to automate the organisation and categorization of files, enhancing data organization and retrieval efficiency for users. Successful implementation of this project will contribute to streamlined file management and improved productivity in various domains.

References

Hoffman., C (2017) 50+ File Extensions that are potentially dangerous on Windows. Available from: https://www.howtogeek.com/137270/50-file-extensions-that-are-potentially-dangerous-on-windows/

Brearcliffe, D.K., Crooks, A. (2021). Creating Intelligent Agents: Combining Agent-Based Modeling with Machine Learning. In: Yang, Z., von Briesen, E. (eds) Proceedings of the 2020 Conference of The Computational Social Science Society of the Americas. Springer Proceedings in Complexity. Springer, Cham. https://doi.org/10.1007/978-3-030-83418-0\_3