Degree Project:

File Handling in Python



Björn Lagerblad

NBI/Handelsakademin

File Handling in Python

2025/02

# 

# Abstract

I chose to write this essay in English so that any future employers or curious visitors to my GitHub can easily understand its purpose.

This project focuses on the development of a Python script to automate file organization using **pathlib, shutil** and **tkinter**.

The script identifies files in a given directory, categorizes them based on file type, and moves them into designated folders for better organization. This automation reduces manual effort and improves overall efficiency in managing files. The script was tested across multiple environments, including Windows, Linux, and Raspberry Pi, ensuring cross-platform compatibility. Future improvements include the addition of a graphical user interface (GUI) to further enhance user experience.

To ensure clarity and correctness, I used ChatGPT to proofread and refine the English wording of this report, as I’m eager to explore new tools and I see it as an excellent way to test the LLM/Human integration of language barriers and terminology.

# Förkortningar och Begrepp

**Skapas automatiskt i Word genom att gå till Referenser > Innehållsförteckning.**

Table of Contents

[1 Abstract 2](#_Toc191406224)

[2 Förkortningar och Begrepp 3](#_Toc191406225)

[3 Introduction 1](#_Toc191406226)

[4 Theory 2](#_Toc191406227)

[4.1 Cross-Platform File Systems 2](#_Toc191406228)

[4.2 Python’s File Management Approach 2](#_Toc191406229)

[4.3 GUI Theory and Principles 3](#_Toc191406230)

[4.4 Error Handling & Robustness 3](#_Toc191406231)

[4.5 Automation & Productivity 3](#_Toc191406232)

[4.6 Summary 4](#_Toc191406233)

[2 Metod 5](#_Toc191406234)

[3 Resultat och Diskussion 6](#_Toc191406235)

[4 Slutsatser 7](#_Toc191406236)

[Appendix A 8](#_Toc191406237)

[Källförteckning 9](#_Toc191406238)

# Introduction

I found it prudent to create something simpler yet with a new challenge: a file organizer that works across different operating systems and includes a functional GUI.

During larger projects, I noticed that many of my classmates have massive download folders with scattered files, resulting in an unorganized mess that makes efficiency worse.

Automating this process can (hopefully) help future students keep their files in order, and it will also serve as a good example of thorough code documentation.

This project focuses on developing a Python script to automate file organization. The main goals of the script are to:

* Identify files in a folder
* Sort them based on file type (e.g., images, documents, videos)
* Move them to specific directories for better organization
* Implement a working GUI with robust error handling
* Ensure compatibility across multiple operating systems

The script is built using **Python 3.11** and its standard libraries, particularly **pathlib** for file handling. If time permits, a **Graphical User Interface (GUI)** using **tkinter** or another Python framework will be added to improve usability.

Ultimately, the aim is to challenge myself with something new (GUI handling in this case) in Python while documenting every step of the process, from planning and coding to testing and analysis.

# Theory

In this section, I’ll explain the main ideas behind how this project will work. First, we’ll look at how different operating systems handle files and why that matters between the different Operating Systems today. Then, I’ll discuss Python’s built-in tools—like **pathlib** and **shutil**—that works such as moving, copying, and deleting files without having to worry about path separators or manual string manipulation. Finally, I’ll give a quick overview of how a GUI (graphical user interface) deals with user interactions, making the program more accessible for people who don’t know terminal commands.

## Cross Platform File Systems

Operating systems like Windows, Linux, and macOS (to name the most popular ones) differ in how they manage file paths, directory structures, and permissions some examples are.

* **Windows:** Commonly uses backslashes(\) in file paths (e.g. C:\Users\Documents), and file names are often case-insensitive.
* **Linux/macOS:** Use forward slashes (/) in file paths (e.g., /home/user/Documents), and typically have case-sensitive file systems.
* **Permissions:** Each system has its own approach to user permissions, read/write access, and hidden files.

For a file organizer to be truly cross-platform, it must handle these variables perfect, or we will get more errors than a new computer in a senior home. By researching libraries and patterns that works away these differences, we can find that some libraries in python will function with that goal in mind.

## Python’s File Management Approach

Python provides multiple ways to interact with files and directories. Historically, the os module (and os path) was the standard tool, meaning you had to manually handle paths in your code. However, modern Python has introduced more user-friendly options that simplify these tasks.:

* **pathlib:**  
  Offers an object-oriented approach to file paths, simplifying code and improving readability. It automatically manages path separators, so the same code works on Windows, Linux, and macOS without manual string manipulation. A great tool with our cross platform goal in mind.
* **shutil:**  
  Provides high-level file operations such as copying, moving, and deleting directories. It complements pathlib by handling tasks that go beyond simple path manipulation, aka we have the tool that works in tandem with pathlib.

## GUI Theory and Principles

A Graphical User Interface (GUI) allows users to interact with an application through visual elements like buttons, menus, and dialog boxes, rather than the command line. In Python, **tkinter** is a common library for building simple GUIs:

* **The User Interface:**
* GUIs rely on events (e.g buttons, key presses and so on from user input) something I have not worked on before and i am approaching this challenge with an open mind.
* **User-Friendliness:**  
  A GUI needs to be as straightforward as possible, so I plan to test it with friends who have minimal technical experience some who admit they struggle even with microwaves. If they can navigate the interface easily, then it’s truly user-friendly.

## Error Handling & Robustness

When working with file operations, robust error handling is essential to prevent data loss or application crashes. Potential issues include:

* **Permission Errors:**  
  Attempts to move or delete files without the necessary read/write permissions.
* **Locked Files:**  
  Some files may be open in other programs, preventing modifications.
* **Invalid Paths:**  
  Missing or broken directory references, especially on systems where a path may not exist or has changed.
* **Different OS Handling:**  
  Different Operating Systems have different sets of rules for file permissions, hidden files and so on, I think this will be the biggest hurdle to get over.

By using exception handling and meaningful error messages, I can help users from clearly know what the issue is so they know what the problem is.

## Automation & Productivity

Manual file sorting can be tedious and prone to human error, especially in environments with high file turnover (e.g. large download folders for teaching a neural network). Automating these processes will in turn:

* **Save Time:**  
  Routinely using this program will save a lot of headaches in not needing to manual sort the files.
* **Reduces Errors:**  
  Consistently using the same script will decrease human error and minimize misplaced filkes.
* **Improves Organization:**  
  Folders and subfolders stay orderly, making it easier to locate and manage files later.

## Summary

By learning about cross-platform file systems, exploring Python’s powerful libraries, and planning a user-friendly GUI with a rudimentary error handling, I’ve laid out a clear foundation for this file organizer. The concepts I covered ranging from path abstractions with pathlib and shutil to the benefits of automation and with what I need to have in mind starting the project.

# Method

I approached this project in several iterative steps, allowing me to gradually build up from basic concepts to a fully functioning file organizer, I’ve documented my journey in bulletin points here.

## Initial Setup

I started by setting up a Git repository to ensure that every change was tracked or if I needed to make a rollback. I created a virtual environment (venv) using Python to isolate my project dependencies from the rest of my system, a good rule when especially when it comes to working in Python. Once the environment was ready, I installed the necessary libraries and froze my requirements in a **requirements.txt** file, a crucial step if I need to reinstall something later down the line or share it with others.

## Research and Warm-Up Exercises

Before diving into the full implementation, I needed to deepen my understanding and experience with our main libraries **pathlib** and **shutil**. Although the official documentation for these libraries is robust, I felt it better to ask GPT for some warm-up exercises. These exercises was simple, such as moving a file or creating a directory, which helped me gain confidence in using these modules.

For instance, I practiced by writing small snippets to:

* Read and list all files in a directory.
* Create a new folder using Path.mkdir().
* Move files from one directory to another using shutil.move().

With these warmup exercises I felt more comfortable to start the project in full.

## Designing the Workflow

Once I felt comfortable with the basics, I moved on to designing the overall workflow for the file organizer. I began by drafting pseudocode to outline my ideas. My initial pseudocode included steps for:

* Identifying files in a target directory.
* Determining each file’s type by checking its extension.
* Sorting the files into corresponding folders (e.g., Images, Documents, Videos).
* Handling files with unknown extensions by placing them into a “Misc” folder.
* Eventually, incorporating a GUI that allows users to select folders more easily.

Since I am less experienced with GUI development, that part of the project was only roughly outlined at this stage, with the main focus being on the file sorting logic.

## Iterative Development and Refinement

Armed with my pseudocode and more caffeine in my system that would probably kill a smaller elephant, I implemented a basic version of the file organizer. In the first iteration, the script successfully moved files into pre-created folders. This version was purely based on command lines in the terminal but helped me verify that the core worked as expected.

After confirming that the script could sort files correctly, I modified the sorting by modifying the folder setup process. Instead of relying on pre-existing folders, I updated the script to automatically create the necessary folders if they didn’t exist and give a print if they were already created.

This change made the program more user friendly and autonomous, reducing the manual setup required before running the script.

I also incorporated exception handling to deal with common issues:

* **Permission Errors:** If a file could not be moved due to insufficient permissions, the script now catches this exception and prints an informative error message.
* **Duplicate Files:** The script checks if a file with the same name already exists in the target folder and skips the move operation to prevent overwriting.
* **Skipping Critical Files:** To ensure that the script itself and essential files (like a requirements file) are not moved, I added specific conditions to skip them.

Throughout the development process, I made frequent commits to Git, committing each significant change and pushing them to my main branch.

## Testing and Feedback

After implementing the core functionality, I conducted extensive testing:

* I created a test directory filled with a variety of file types, including images, documents, videos, and miscellaneous files.
* I introduced dummy cases, such as files without extensions, duplicate files, and files that were open in other applications, to see how the script handled these situations.
* I also tested the script on different operating systems (Windows, Linux, and Raspberry Pi) to verify cross platform compatibility.
* As I am still a junior and not always the best at commenting my own code I’ve asked GPT to help me with comments on my code as to not confuse future visitors on my program.

Feedback from these tests were a learning experience. I ran the script in a controlled environment and checked my console output to ensure that each action (e.g. folder creation, file movement, error handling) was logged clearly and that no unseen errors would come to fruition. This step confirmed that the script behaved as expected and helped me identify areas for further improvement.

## Future Enhancements after the First Stage

Although the current version of the file organizer is functional and meets the primary goal of sorting files, I have identified several areas for future development:

* **GUI Integration:** I plan to develop a user-friendly interface using **tkinter** to allow users to select directories and view progress in real time.
* **Advanced Sorting Rules:** In addition to sorting by file extension, I would like to implement more sophisticated categorization (e.g. grouping files by date or size).
* **Secure a new caffeine supplier:** My caffeine reserves have reached dangerously low levels, and immediate resupply is critical.

## Summary of the First Stage

In summary, my method in the first stage of development involved a structured, iterative approach:

* Setting up the project environment with Git, a virtual environment, and dependency management.
* Researching and practicing basic file operations with **pathlib** and **shutil** through warmup exercises.
* Drafting and refining pseudocode to outline the file organizer’s workflow.
* Implementing the script in iterative stages, beginning with basic file movement and progressing to automated folder creation and a morerobust handling.
* Testing the solution across different platforms and gathering feedback to guide further enhancements.

This approach ensures that the final product is not only functional but also well-documented, maintainable, and prepared for future upgrades such as GUI integration.

# Resultat och Diskussion

|  |  |
| --- | --- |
| **RMSE för olika modeller** | |
| Enkel Linjär Regression | xx |
| Lasso | xx |
| Ridge | xx |

Tabell 1: Root Mean Squared Error (RMSE) för de fyra valda modellerna.

En bild som visar text, Teckensnitt, skärmbild, linje

Automatiskt genererad beskrivning

Figur : Hur man lägger in tabell eller figur nummer samt beskrivning.

# Slutsatser

Här besvarar du frågeställningarna.

# Appendix A

# Källförteckning