

A Client-Side App Model for Classifying and Storing Documents

Bahaeldein Elhussein¹

College of Computing and Information Technology
University of Bisha, Bisha, Saudi Arabia
bmelamin@ub.edu.sa

Abdelrahman Elsharif Karrar³

College of Computer Science and Engineering
Taibah University, Medina, Saudi Arabia
akarrar@taibahu.edu.sa

Mahmoud Khalifa²

College of Computing and Information Technology University
of Bisha, Bisha, Saudi Arabia
mkhalifa@ub.edu.sa

Mohammed Mujib Alsharani⁴

College of Computing and Information Technology University
of Bisha, Bisha, Saudi Arabia
mmalshahrani@ub.edu.sa

Abstract

Due to the large number of documents that are important to people and many of their requests from time to time to perform an essential official procedure, this requires a practical arrangement and organization for them. When necessary, many people struggle with effectively arranging official documents that enable display, which takes a lot of time and effort. Also, no mobile apps specialize in professionally preserving essential electronic records and displaying them when needed. Dataset consisting of 10,841 rows and 13 columns was analyzed using Anaconda, Python, and Mito Data Science new tool obtained from Google Play. The research was conducted using the quantitative descriptive approach. The presented solution is a model specialized in saving essential documents, categorizing according to the user's desire, and displaying them when needed. It is possible to send in an image or a pdf file. Aside from identifying file kinds like PDFs and pictures, the model also looks for and verifies specific file extensions. The file extension and its properties are checked before sharing or saving it by applying the similarity algorithm (Levenshtein). Our method effectively and efficiently facilitated the search process, saving the user time and effort. In conclusion, such an application is not available, which facilitates the process of classifying documents effectively and displaying them quickly and easily for people for printing or sending to some official procedures, and it is considered one of the applications that greatly help in preserving time, effort, and money for people.

Keywords:

Classifying Official Documents, Storing Official Documents, Show Images and PDF Files, Mobile App

1. Introduction

Since the mobile has become today, it dispenses with other purposes, and as an example for clarification of travel, the mobile is sufficient only. However, it is essential to provide documents in photos or files such as passports, official documents such as cards, and other important papers that help accomplish various

tasks. Through research, we found that it is not available. An application that stores and categorizes official documents in an easy-to-use manner so that the user may access and submit them as necessary for official procedures.

Technology and the World Wide Web have entered almost every field and sector of life over the last two decades [1]. The world is developing and becoming increasingly digital. Electronic papers are becoming increasingly popular with traditional paper-based documents [2]. Text classification is a way of finding and classifying large amounts of material for easy reference. Pages on the Internet, books, journals, and social media have become rich knowledge sources that must be categorized and organized for quick access [3].

Hence the importance of categorizing official documents and making a mobile application dedicated to this service, as an example of a personal wallet to classify and save official papers in the form of an electronic document referred to when needed. The fast growth of electronic records has resulted in issues such as unstructured data, which necessitates more time and effort to find important papers. Text Document Classification (TDC) is essential for information processing and retrieval since it categorizes unstructured documents into specified classifications [4]. Everything is going paperless in today's competitive environment [5]. Electronic document management solutions are now commonplace in many businesses [6].

Mobile devices are an assertive economic, mass communication, educational instrument, and communication Tool. Due to substantial expansion, mobile devices and the Internet are increasingly

employed in learning and education [7]. Thus, the mobile has become involved in many things, facilitating the transmission and communication process, and reducing a lot of effort, money, and time for users, hence creating an application that is interested in official documents for people. In everyday life, identity verification methods are much employed. These systems rely on official documents holding personal information (i.e., passports, ID cards, and driving licenses) [8]. One problem that significantly affects wasting a lot of time, effort, and money is large numbers of official and unofficial documents. Moreover, there is a lack of a system dedicated to arranging and organizing them, which we find in many electronic devices. Many people search for these documents using the traditional search methods, and sometimes they do not get these important documents using the standard search methods.

One of the objectives of submitting the paper is to arrange and organize the official documents, for example, passports, cards, licenses, and essential documents, in an electronic wallet on the mobile, which enables the user to know them and display them or send them when needed, especially when required to make official procedures.

Transactions. One of the objectives of the presented program is to preserve the effort, time, and money of the users of the program by classifying and arranging documents in a standard and simplified manner that enables the user to obtain them when needed. From those, we conclude the importance of presenting a program concerned with arranging and preserving official documents and submitting them or sending them when required, as the model presented in this paper. Most governments now recognize the necessity of electronic government (e-government) for modern public administration [9]. We also find it supports the idea of the system working to save and organize official documents.

The classification of papers has proven difficult in recent years. There is no possible mechanism or systematic process for assigning appropriate labels to many documents in document classification. As a result, a method for classifying PDF documents that are both efficient and accurate is required [10]. The amount of data produced worldwide has increased dramatically during the last two decades [11]. Thus, arranging and organizing information has become one of the basic requirements in various fields. Society has entered the digital transformation era due to the rapid development of digital technology [12]. Document categorization is the process of dividing a collection of documents into

different, highly related classes or groups depending on the content of the documents [13]. The documents are organized in a way that helps the user to obtain the file or files quickly and conveniently in the provided program. The increasing growth of electronic documents has resulted in issues such as unstructured data, which necessitates more time and effort to find a relevant document. Text Document Classification (TDC) is essential for information processing and retrieval since it organizes Unstructured documents into predetermined classifications [14].

One of the problems with cards and important documents is their large number, and our electronics are not well arranged, enabling the user to retrieve them to display or send them quickly. Of course, it is possible to set them in hard copy paper copies within a file or the like. However, in the case of a soft copy, we rarely get an application. They are interested in arranging and organizing documents professionally and retrieving them when needed for sending them in the form of a copy of an image or a PDF. This program is very concerned with arranging and organizing essential documents straightforwardly and displaying or sending them when needed. The electronic wallet keeps cards and documents categorized and tidy by taking a picture of them and saving them. We notice that most of the programs used do not care about arranging and organizing documents clearly and thus will need to search using the search engine, which wastes much time. The program is interested in organizing cards and Documents, thus significantly saving much time, effort, and money. There are many features, such as facilitating sharing between other programs, such as various social networking programs. Important documents these days are required from time to time and, in many cases, requested in the form of a digital copy. We often face problems with saving them in the form of digital images in an organized manner that facilitates displaying them when needed. Therefore, this model was introduced, which helps arrange and organize digital documents and display or send them when required. Thus, the model helps reduce time, effort, and money for people. The file extension and its properties are checked before sharing or saving it by applying the similarity algorithm (Levenshtein). One of the objectives of the mobile application is to organize and arrange essential cards and documents to facilitate the process of displaying and sending them when needed in PDF or image format, thus saving much time, effort, and money for the user. The application saves critical official documents and categorizes them according to the user's desire, illustrating the idea.

Another, he needs a copy of it to attach it to carry out official transactions.

Moreover, many people save essential documents on a computer or laptop or are unorganized on a mobile device. Random method and thus consumes much time to retrieve them or display them to perform some official tasks that require official papers. Find that the file helps greatly in arranging and organizing Official documents effectively and showing them when needed, which helps effectively save time and effort for the user.

This research paper has been organized into sections, including a literature review in Section II and the importance of transformation and development with technology; section III contains the research methodology and the data obtained from Google Play, analyzed and extracted results from them; the results that supported the presented research were explained and clarified in Section IV: Discussion and Conclusion, the discussion of the presented research and its conclusion, Finally, Section VI the obtained results and their discussion, A recommendation was made about the submitted paper and what could be presented in the future related to the research presented.

II. RELATED WORK

According to [15] and [16]. Organizations rely on these records to achieve their goals and extract knowledge for decision-making and problem-solving. As a result, these records should be kept to prove what operations were carried out. To reduce the danger of fraud and corruption in the administration of information created in the event of legal disputes and minimize the potential for legal issues, these records should be managed methodically and effectively.

According to [17], The benefits of employing automated data input from scanned copies of contract department papers over human input are discussed. It is proposed to apply machine learning, provided by a neural network, to improve data transfer from a scanned document to a corporate database management system in the energy industry. Machine learning allows data classification for examined documents, ensuring the proper template selection when creating an electronic record. The techniques used to construct a software module for data extraction are justified, and the operation principle is presented.

This paper compared the scanned data entry with the manually entered data. With this enhancement, data from scanned documents will be transferred more

quickly to the database. The machine language is used for the documents from which the information was obtained, and thus it is possible to use the correct model when creating the electronic document. Compared with the scientific paper presented here, we find a difference in that the presented program is dedicated to displaying classified documents, official documents, or classified documents to send or view.

Another research paper [18] defined an electronic document management system as a set of tools that manages numerous forms of document usage, storage, and creation of documents throughout an organization (EDMS). Document management is a centralized software system for managing and capturing scanned documents' images and digital files. Electronic document management systems and enterprise content management (ECM) systems share similar capabilities; nevertheless, the document management software system.

Innovative card user behavior classification in a study by [19] is critical in public transit demand analysis. It enables people to comprehend the chronology of their activities across time. However, traditional measures such as Euclidean distance are ineffective when dealing with time-series classification. This article presents a method for classifying public transit smart card users' daily transactions, recorded as a time series, to tackle this challenge. To decipher the clients' transient examples, the picked approach utilizes cross-connection distance (CCD), various leveled groupings, and subgroups by estimated quality. The distance between unique worldly twisting (DTW) and the bunching results is thought about (a standard technique to quantify time-series distance). An R application is created to evaluate the strategy on an actual dataset of intelligent card data transfers after a brief pedagogical example to demonstrate the DTW and CCD notions. The information in this dataset relates to Gatineau's public transportation usage in September 2013. The findings show that CCD Outperforms ZDTW in classifying time series and that the classification approach distinguishes between different public transit customers' everyday routines. The findings will aid transit authorities in providing better services to intelligent card users from various demographics.

The use of data and the optimization of structured and active documents, such as emails and other defined forms, PDF files, PowerPoint, and Word ECM frameworks, monitor rich media types and manage

unstructured content in addition to documents and Excel spreadsheets [20]. The research above analyzed the smart card for public transportation to understand the chronology of people's activities over time. It is possible to save a picture of such cards and categorize them according to their name, in this case, by mean a user. Indeed, this method helps save many important documents and cards and classify them in an intelligent way that facilitates displaying them when needed.

According to [21], Paper documents, such as birth certificates, death certificates, identity cards, and official documents, play an essential role in our daily lives. However, as digital technology advances, concerns regarding the legitimacy of these paper papers have grown. This paper presents a method for determining the age of a document using Local Binary Pattern (LBP) features. The essential source of information is paper documents, which have been used for communication, record keeping, and proofing. A paper document is a piece of paper used for writing or printing. In today's world, we use a significant number of paper documents. Watermarking, embed signatures, and printed patterns have been used; nevertheless, various difficulties linked to document security have arisen due to the misuse of digital technology. Document age identification is critical in verifying a document's authenticity and originality to resolve this problem. This paper contains 500 printed items that were published throughout the year. The above image is concerned with knowing the document's age, and thus the paper is also worried about saving documents, understanding them, and checking them. Compared to the research presented in this paper with the analysis above, we find differences, including that the presented research is concerned with official documents such as certificates, cards, licenses, passports, invoices that need to be saved and arranged to display in the future, it will not be sent as an image or as a file.

A study [22] mentioned that documents play a critical role in our daily lives. Document forgery can result in financial losses, misjudgments, and a loss of respect and goodwill. As a result, papers must be safeguarded against threats such as forgery, fabrication, and tempering, and their authenticity must be easily verified. Modern technology has numerous recognized approaches for validating authenticity and integrity, such as blockchain and digital signatures. Due to their complexity, high cost, and implementation challenges, most methods are not instantly suited for public usage, and a simple verification method is still unavailable to

the public. As a result, this paper offers a document verification method that intends to provide authenticity, (ii) integrity, (iii) availability, and (iv) non-repudiation. Because the proposed method has no licensing fees, is simple to implement, and can be used for electronic and printed publications, it will benefit the public. It is worth noting that the proposed method will allow you to verify the document's authenticity using only your smartphone quickly. We find that this paper is concerned with documents and how they are presented, and to be more on the security side of the documents and verify their authenticity; on the other hand, The paper presented in this research is concerned with arranging documents and keeping them coordinated and organized.

III. Methodology

A. ANALYSIS DATASET

Dataset consisting of 10841 rows and 13 columns was analyzed, taken from Google Play Store, from the date of the month of seven days 31 of the year 2013 to six days seven days 2018.

Deep learning is a valuable way to deal with BIG Data and boost identification performance. It has recently been successfully used in various applications [23]. The research was conducted using a quantitative descriptive method. The data collection was clustered and partitioned into subclasses containing related items using the clustering technique. It was a similarity-based data partitioning [24]. Clustering mined the data set using a partitioning method (data partition). Analyzed the data using Python and the Anaconda Mito tool.

In figure 1. the data size shows the number of quantitative rows and columns displayed on the right-hand side of the analysis result. From the exact Figure, we find the name of the programs, their evaluation, the number of residents of the program, the program's size, the number of downloads for each program, and finally, the type of the program, if it was free or paid.

App	Category	Rating	Reviews	Size	Installs	Type
Photo Editor & Candy Camera & Grid & Scrapbook	ART_AND_DESIGN	4.1	159	190	10,000+	Free
Coloring book means	ART_AND_DESIGN	3.9	967	140	500,000+	Free
U Launcher Lite - FREE Live Cool Themes, Hide Apps	ART_AND_DESIGN	4.7	87510	8.7M	5,000,000+	Free
Sketch - Draw & Paint	ART_AND_DESIGN	4.5	215644	25M	50,000,000+	Free
Pixel Draw - Number Art Coloring Book	ART_AND_DESIGN	4.3	967	2.8M	100,000+	Free
Paper Flowers Instructions	ART_AND_DESIGN	4.4	167	5.6M	50,000+	Free
Smoke Effect Photo Maker - Smoke Editor	ART_AND_DESIGN	3.8	178	190	50,000+	Free
Infuse Painter	ART_AND_DESIGN	4.1	36815	290	1,000,000+	Free
Garden Coloring Book	ART_AND_DESIGN	4.4	13791	330	1,000,000+	Free
Kids Paint Free - Drawing Fun	ART_AND_DESIGN	4.7	121	3.1M	10,000+	Free
Text on Photo - Fontface	ART_AND_DESIGN	4.4	13802	280	1,000,000+	Free
Name Art Photo Editor - Focus & Filters	ART_AND_DESIGN	4.4	8788	120	1,000,000+	Free
Letter Name On My Photo Editor	ART_AND_DESIGN	4.2	44829	280	10,000,000+	Free
Mandala Coloring Book	ART_AND_DESIGN	4.6	4326	210	100,000+	Free

Fig. 1. Shows The Number of 10841 Rows and 13 Columns in TheDataset.

After analyzing and classifying electronic wallets, about 14 were found, most of which specialized in financial transactions and shopping. The search and classification were carried out with other words such as documents, and the result was the presence of two applications specialized in scanning documents. They were obtaining a program specialized in saving and organizing important documents and cards in an orderly and organized manner, displaying and sending them whenever needed.

The analysis found that most of the programs are free and not paid, about 10,029 programs are free, and the rest of the other programs are paid, approximately 800 programs. Through this, the program must be accessible to market it more professionally, and downloading it helps immensely. In the marketing process for it, one of the critical points that have been focused on is the program's effectiveness in arranging documents clearly and correctly. The graph compares the amount of free and paid programs, which is a vast difference. Before that, empty fields were filtered, and zero values were highlighted.

Figure. 2. shows the level of the free program compared to the paid programs. From this, we conclude that the percentage of downloaded free programs is more than the paid programs and that the designers of free programs have other sources of income from these programs, such as advertisements, for example. We benefit from the importance of the program being free with a good possibility. That is, it benefits the user. Whenever the program is free and required, meaning that the user needs it, this is a feature that distinguishes it and enables it to spread and expand with time. We can know the pros and cons by analyzing the use results over time, the more prepared the users. As a result, we can analyze the data and extract results that may be used to enhance and develop the software and necessary changes continually. Therefore, we decided to make the application accessible to study user behavior and continuous improvement of the program, and the quantitative result of the analysis encourages that quantitatively. It is shown in the Figure below.

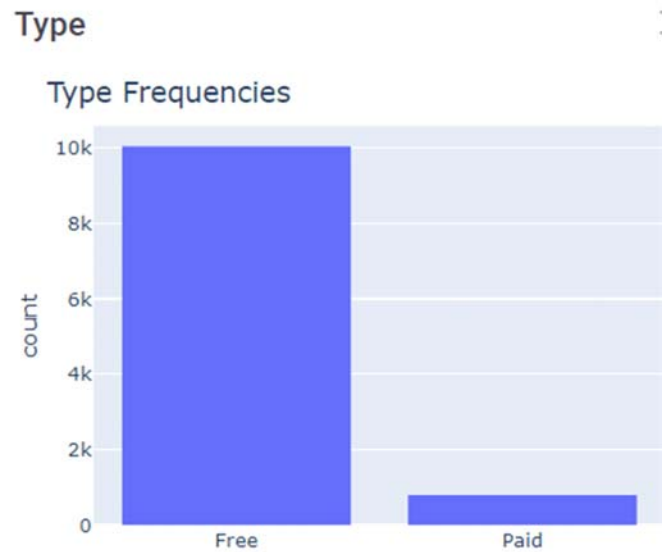


Fig. 2. Shows Free and Paid Programs Levels in Dataset.

Figure. 3. depicts that most of the programs in the analyzed database are devoted to family, communication, lifestyle, and domestic life. It is non-existent. It is considered an incentive to create a program dedicated to such essential operations. Yes, many programs help keep documents, but not in a clear and specific way; that is, the need to search and dig inside folders or the device to get the official document, which is considered random.

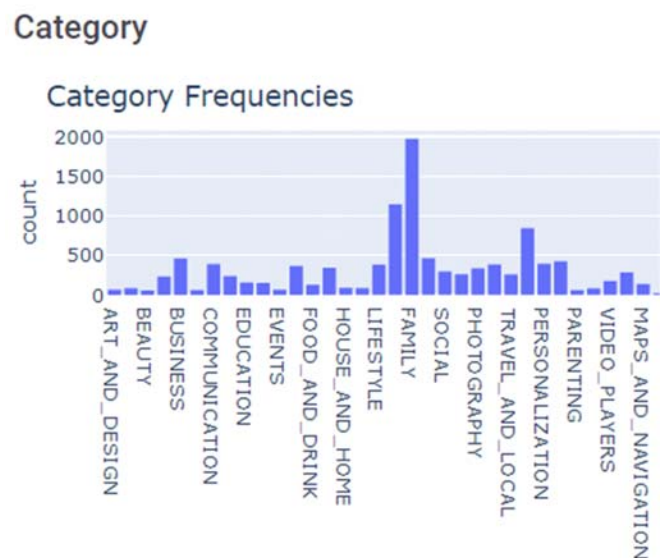


Fig. 3. Shows Most of The Programs in The Data Set.

Figure. 4. briefly shows the volume of data analyzed; the only programs that take up much space or are used extensively by users are the family programs.

Column Summary Statistics

count	10841
unique	34
top	FAMILY
freq	1972
count: NaN	0

Figure. 4. Show The Data Set Column Summary Statistics.

Here, the programs were arranged in ascending order, and then the programs were arranged in descending order. Through this analysis, it became clear that document management and classification programs are not well available, and through research, exploration, and analysis, no such programs were found in the database. Such applications must be given attention and developed due to this finding.

TM was matching the analysis, x neon vector the number of programs and their percentage, the first of which is family similarities wwrite72 programs ago within a stage of categories 18%, games with several 1144 programs, percentage 5.10%, tools with several 843 programs ad percentage of 778% and the rest of the programs it number and percentages shown in the following analysis result.

In Figure. 5. the data were categorized descending, i.e., from the highest programs in use to the minor programs downloaded and used. Using this method, we knew many, including that the programs used and downloaded in a significant way are news programs and magazines, and their quantitative number is shown in the Figure below. Two hundred and eighty-three programs and their percentage within the database is two point six, followed by shopping programs, which number two hundred and sixty programs and their percentage is two point four, then travel and travel programs with their number and percentage, quantitatively illustrates the Figure No. five that. Through the analysis, we found that the programs allocated to official documents are sure to be non-existent. They do not exist, so it was considered

an encouraging result to work on implementing the research and moving forward with it.

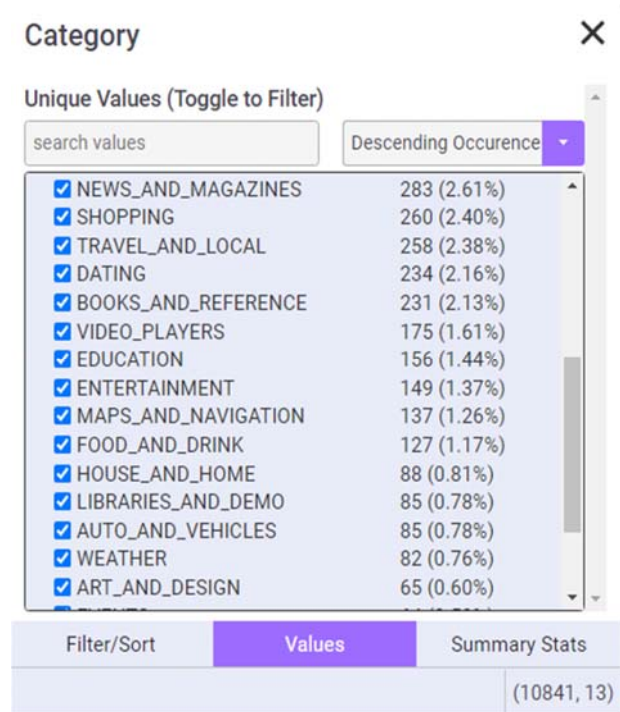


Fig.5. Shows The Programs Arranged in Descending Order with Their Numbers and Percentage in the Dataset.

B. Classification

An online document classification that depends on authentication bases for files maintains reliable e-commerce and e-government applications. Authentication of digitally capture document images from hardcopy documents is an essential forensic task that attracts much attention [23].

Our classification procedure matches the extension of any attached files against the stored lexical of files signature feature using the Levenshtein algorithm Similarities -(shown in equation1) -. Moreover, the detected output differentiates between the image class (png, jpg) and PDF (pdf) class. The system must ensure that no attached files in any other type are stored except the image class or PDF class in the directory.

$$ev(a,b)(i,j) = \begin{cases} \max(i,j) \\ \min \left\{ \begin{array}{l} Lev_{a,b}(i-1,j) + 1 \\ Lev_{a,b}(i,j-1) + 1 \\ Lev_{a,b}(i-j-1) + 1_{a_i \neq b_j} \end{array} \right\} \end{cases}$$

(1) Levenshtein algorithm Similarities

C. The Introduced Model

The algorithm used to create a directory (a new folder with the proper name depends on filing an extension). The diagram below shows a simplified program working through the presented system. First, the user deals with the program by saving and categorizing documents according to his desire. For example, the user can create a folder dedicated to preserving important documents. They are processed and classified, images or files, and kept after processing, classification, and arrangement. Also, according to the user's desire, he can retrieve them from databases and display them to send or print. Additionally, Documents are also saved in the date, day, and time to facilitate knowing the information accurately when needed. Today, it is possible to benefit from various companies worldwide on the Internet [24]. Furthermore, we find the importance of having official documents within the user's reach, coordinated, and arranged. These institutions may regularly require some official documents or information for an official procedure or a necessary procedure. Thus, we find that the provided application helps facilitate the tasks for the user in a good way.

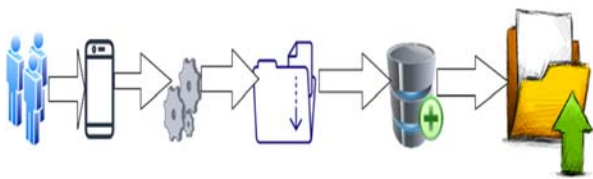


Figure. 6. Shows The Model Introduced.

IV. EXPERIMENTAL RESULT

Compared with the various applications that are concerned with saving documents, we did not find an application that specialized in saving official documents such as passports and actual receipts, in addition to certificates such as university certificates and to all the documents that the user needs from time to time for an official procedure. From here, we find the importance of the application. By comparison, With some applications that are concerned with government documents and papers, we found that, for example, but not limited to, the majority of programs interested in scanning documents and sending them, such as Cam Scanner and Adobe applications, to scan files, and from this, we found that the majority of applications are concerned with scanning operations for documents and neglect the aspect of classifying and organizing files in a way that helps the user to find them. Efficiently and quickly, in case you need it to print or send it to a specific official procedure, the program is concerned with saving and classifying

official documents for the future. It is easy to obtain the document and send it when needed.

My experience with the application showed effectiveness in classifying and coordinating documents and saving them by enabling the user to retrieve and display them quickly and conveniently. One of the points mentioned here is the continuous improvement of the application to facilitate saving and classifying documents, especially official documents.

The model's significance is clear from the experiments and illumination research. Among the reasons that confirm the correctness of the information provided above is the importance of documents for people, such as certificates, receipts, cards, family passports, necessary commercial contracts that need to be preserved electronically, as well as land contracts and various projects that people need after a long period has passed for them or no copy of them.

Utilizing tools such as Anaconda, Python, and Mito Data Science to analyze a dataset of 10,841 rows and 13 columns collected from Google Play, we conducted our research in this work. A similar Java software algorithm was then used to classify added files to the appropriate folders category. When it comes to exhibiting these papers for viewing, sending, or printing, the program has proven to be an efficient and intelligent investment through real-world experience for various official or unofficial procedures. As a result, it was saved in an organized and neat manner on the mobile device. Analysis of the survey data obtained in this project shows that the unpaid applications have a considerable percentage of downloads compared to the paid applications, as explained above in the analysis. The widely used programs are family, personal, and programs that care about financial matters. Analyzing the results There are not any other applications like the one you are looking at, and that is why the analysis in the analysis section shows us that the application you are looking at is one of those that is commonly used, so it is a metaphor for the one you are looking at.

V. DISCUSSION AND CONCLUSION

Most of the published scientific papers did not address the process of arranging and classifying documents according to their importance.

By looking at the different research published previously, there are many differences with the research presented, including that the research is concerned with arranging, organizing, and classifying the essential documents that we need from time to time, such as cards, passports, and

various documents, in a way that enables them to be retrieved and displayed readily and conveniently when need. The program is a mobile application and thus contributes significantly to saving much time, effort, and money. When these official documents are needed to complete some tasks, the program dramatically helps.

For clarification, we sometimes need to do some official procedures, and here some copies of official documents are needed to complete the official procedure. It was discovered that the presented application efficiently aids in this because documents are classified distinctively and are familiar to the user. Thus, he can retrieve them, view them, and send them when needed.

Finally, such programs need attention and development and are effective for many people.

Furthermore, it has many positive aspects, and there are many applications concerned with converting documents from one form to another, for example, from image to PDF and scanner programs. With time these programs are crowded with documents that are not arranged organized effectively. Some of the documents are important. Others are not, and here Such provided program maintains important documents and arranges them, which helps accomplish many tasks for users of the program.

vii. RECOMMENDATION (FUTURE WORK)

Some points must be mentioned to propose the presented application and its future development because of the considerable progress in information technology and its practical support in saving time, effort, and money for different people and institutions every day. Due to speed, the module needs continuous development and research to provide its services more effectively. The application is essential, especially for people who deal with papers and documents on an ongoing basis and need them from time to time. They need copies of them to send or print by presenting the performance of some Formal or informal transactions. Through the application and analysis, we find that the provided module dramatically helps maintain documents more effectively and organized and thus facilitates obtaining them from the mobile application. Since many people rely on these programs, we urge that they continue to be developed and maintained in the future. Even if programs like this one were not discovered through analysis, it serves as an analogy because the provided application is one of those indicated above that are needed for general usage, as the analysis in the analysis section shows."

We recommend the continuous improvement of the application. The reason is that the program is essential in maintaining the official papers for the user on the mobile. One of the advantages of the mobile today is that the user

can travel and move using the mobile-only without the need to take other purposes to move between many countries. From here, we found the importance of the program. The application that helps and relieves many of the symptoms for the person on the move from one place to another, all the official user documents are present with him and saved in a coordinated and classified manner that enables him to display them quickly and conveniently when he needs them. Here we must mention an important point also, which is that the continuous improvement of the application is one of the points where The task for the application is beneficial in the development of the system and its effective fulfillment of the purpose for which it was created over time. It saves much time, effort, and money for the user.

REFERENCES

- [1] B. A. Hendal, "Kuwait University faculty's use of electronic resources during the COVID-19 pandemic," *Digital Library Perspectives*, vol. 36, no. 3, pp. 429-436, 2020.
- [2] D. Prasad, A. Gadpal, K. Kapadni, M. Visave, K. Sultanpure, M. Visave, and K. Sultanpure, "CascadeTabNet: An approach for an end to end table detection and structure recognition from image-based documents," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020.
- [3] A. Wahdan, S. A. Hantoobi, S. A. Salloum and K. Shaalan, "A systematic review of text classification research based on deep learning models in Arabic language," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 6, pp. 6629-6643, 2020.
- [4] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood, and M. T. Sadiq, "Document-Level Text Classification Using Single-Layer Multisize Filters Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 42689-42707, 2020.
- [5] M. Kumar, S. Gupta, and N. Mohan, "A computational approach for printed document forensics using SURF and ORB features," *Soft Computing*, vol. 24, no. 1, p. 13197-13208, 2020.
- [6] I. Lvovich, Y. Lvovich, A. Preobrazhenskiy, Y. Preobrazhenskiy, and O. Choporov, "Optimisation of the Subsystem for the Movement of Electronic Documents in Educational Organization," in *2021 1st International Conference on Technology Enhanced Learning in Higher Education*, 2021.
- [7] S. Sophonhiranrak, "Features, barriers, and influencing factors of mobile learning in higher education: A systematic review," *Heliyon*, vol. 7, no. 4, 2021.
- [8] S. Gonzalez, A. Valenzuela, and J. Tapia, "Hybrid Two-Stage Architecture for Tampering Detection of Chipless ID Cards," *IEEE Transactions on Biometrics Behavior and Identity Science*, vol. 3, no. 1, pp. 89-100, 2020.
- [9] A. Ayaz and M. Yanartaş, "An analysis on the unified theory of acceptance and use of technology theory (UTAUT): Acceptance of electronic document management system (EDMS)," *Computers in Human Behavior Reports*, vol. 2, 2020.
- [10] M. B. Divyanshu Singh and V. Yadav, "PDF Classification Using Logistic Regression and Latent Dirichlet Allocation," in *Proceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, 2022.

- [11] P. N. Sawadogo, J. Darmont, and C. Noûs, "Joint Management and Analysis of Textual Documents and Tabular Data Within the AUDAL Data Lake," in *European Conference on Advances in Databases and Information Systems*, 2021.
- [12] S. Chen, "Research on the Strategy of Electronic Documents Management of Universities Based on Digital Campus," in *Proceedings of the 2020 International Conference on Advanced Education, Management and Information Technology*, 2020.
- [13] S. K. Sreedhar, S. Ahmed, P. M. Flora, L. Hemanth, J. Aishwarya and R. G. Naik, "An Improved Approach of Unstructured Text Document Classification Using Predetermined Text Model and Probability Technique," in *Proceedings of the First International Conference on Advanced Scientific Innovation in Science, Engineering and Technology*, Chennai, India, 2021.
- [14] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood and M. T. Sadiq, "Document-Level Text Classification Using Single-Layer Multisize Filters Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 42689-42707, 2020.
- [15] B. Hawash, U. A. Mokhtar, Z. M. Yusof, and M. Mukred, "The adoption of electronic records management system (ERMS) in the Yemeni oil and gas sector: Influencing factors," *Records Management Journal*, vol. 30, no. 1, pp. 1-22, 2020.
- [16] A. E. Karrar, "A Proposed Model for Improving the Performance of Knowledge Bases in Real-World Applications by Extracting Semantic Information," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 2, pp. 116-123, 2022.
- [17] T. Levina, A. Rodionov, and R. Farkhutdinov, "Software module for extracting data from electronic documents," in *2020 International Conference on Electrotechnical Complexes and Systems*, 2020.
- [18] A. Rosa, I. Pustokhina, E. Lydia, K. Shankar, and M. Huda, "Concept of Electronic Document Management System (EDMS) as an Efficient Tool for Storing Document," *Computer Science*, 2019.
- [19] L. He, B. Agard, and M. Trépanier, "A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method," *Transportmetrica A: Transport Science*, vol. 16, no. 2, pp. 56-75, 2020.
- [20] P. Gonasagi, R. Pardeshi, and M. Hangarge, "Classification of Documents based on Local Binary Pattern Feature through Age Analysis," in *Ambient Communications and Computer Systems. Advances in Intelligent Systems and Computing*, Springer, Singapore, 2020, pp. 265-271.
- [21] M. Haque, N. Adnan, M. A. Kabir, M. R. A. Rashid, A. S. M. Yasin, and M. S. Pervez, "An Innovative Approach of Verification Mechanism for Electronic and Printed Documents," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 8, pp. 623-627, 2020.
- [22] D. P. V. Hoai, H.-T. Duong and V. T. Hoang, "Text recognition for Vietnamese identity card based on deep features network," *International Journal on Document Analysis and Recognition*, vol. 24, no. 1-2, pp. 123-131, 2021.
- [23] S. Z. F. L. J. H. Changsheng Chen, "Domain Generalization for Document Authentication against Practical Recapturing Attacks," arXiv:2101.01404, 2021.
- [24] M. Umair, F. Majeed, M. Shoaib, M. Q. Saleem, M. S. Adrees, A. E. Karrar, S. Khurram, M. Shafiq and J.-G. Choi, "Main Path Analysis to Filter Unbiased Literature," *Intelligent Automation and Soft Computing*, vol. 32, no. 2, pp. 1179-1194, 2022.