

Invoice Processing using Robotic Process Automation (UiPath tool) and Artificial Intelligence

Aishwarya Bhargava

ABSTRACT

This paper describes our recent effort to develop an automative, efficient, accurate, and secure application to transform invoice processing in Finance operations. As a prime example of the technology's potential for driving efficiency, Robotic Process Automation (RPA) can be applied to several finance and accounting operations, invoice processing. RPA bot can automate data input, error reconciliation, and some of the decision-making required by finance staff when processing invoices. At the same time, automation can limit errors in such processes and reduce the need for manual exception handling. Most of the invoices are delivered via emails these days, and with the increase in this digital world, the world of scammers has also increased. The finance world is most prone to scams. So, in the paper we will be introducing a front-line screening of mails before they get downloaded for invoice processing, the screening will be done with the help of Artificial Intelligence.

KEYWORDS: Artificial Intelligence, Invoice processing, Machine Learning, Robotic Process Automation, UiPath tool

I. INTRODUCTION

According to various research papers published in the past decades, it can be said that manual invoice processing takes a number of days for complete execution and even after spending a lot of time, the outcome is not an ideal output as it might be prone to human errors. There are multiple drawbacks related to manual invoice processing like missing invoices, confusing invoices, missing data, errors in the invoice data extraction, wrong or missing contact information, and excessive time consumption. In order to eliminate all these errors to some extent, a few years ago the methodology of automated invoice processing was introduced. Many technologies have accomplished a great deal of efficiency and accuracy in this field. In this research paper, we will be discussing the implementation of invoice processing using Robotic Process Automation technology and the implementation will be done using the UiPath tool with the security of Artificial Intelligence.

Due to the advancement in this era of digitization, the dependency on email has been increasing day by day. The increasing dependency calls for a way to manage the huge amount of data or emails. The emails conveyed include important as well as phishing emails. Phishing emails often lead to malicious websites and result in sharing personal details with the attackers. These phishing emails might look like legitimate invoice emails to the company but on the back-end lead to the downfall of the finances of the company. For the safety of the company or any other organization, it needs to deploy the program of Artificial Intelligence which will identify the scams and will not let them move forward for further processing.

II. LITERATURE

[1] Artificial Intelligence

Artificial Intelligence is composed of two words Artificial and Intelligence, where Artificial defines "man-made," and intelligence defines "thinking power", hence AI means "a man-made thinking power." So, we can define AI as: "It is a branch of computer science by which we can create intelligent machines which can behave like a human, think like humans, and able to make decisions." AI is an intelligent entity created by humans. It is capable of performing tasks intelligently without being explicitly instructed. It is also capable of thinking and acting rationally and humanely.

- **Machine Learning:** ML teaches a machine how to make inferences and decisions based on past experience. It identifies patterns, analyses past data to infer the meaning of these data points to reach a possible conclusion without having to involve human experience. This automation to reach conclusions by evaluating data, saves a human time for businesses and helps them make a better decision.
- **Deep Learning:** Deep Learning is an ML technique. It teaches a machine to process inputs through layers in order to classify, infer and predict the outcome.
- **Neural Networks:** Neural Networks work on the similar principles as of Human Neural cells. They are a series of algorithms that captures the relationship between various underlying variables and processes the data as a human brain does.
- **Natural Language Processing:** NLP is a science of reading, understanding, interpreting a language by a machine. Once a machine understands what the user intends to communicate, it responds accordingly.
- **Computer Vision:** Computer vision algorithms tries to understand an image by breaking down an image and studying different parts of the objects. This helps the machine classify and learn from a set of images, to make a better output decision based on previous observations.

- **Cognitive Computing:** Cognitive computing algorithms try to mimic a human brain by analyzing text/speech/images/objects in a manner that a human does and tries to give the desired output.
- [2] Advantages of using AI
- AI drives down the time taken to perform a task. It enables multi-tasking and eases the workload for existing resources.
 - AI enables the execution of hitherto complex tasks without significant cost outlays.
 - AI operates 24x7 without interruption or breaks and has no downtime
 - AI augments the capabilities of differently abled individuals
 - AI has mass market potential; it can be deployed across industries.
 - AI facilitates decision-making by making the process faster and smarter.
- [3] Disadvantages of using AI
- The implementation cost of AI is very high.
 - The difficulties with software development for AI implementation are that the development of software is slow and expensive. Few efficient programmers are available to develop software to implement artificial intelligence.
 - A robot is one of the implementations of Artificial intelligence with them replacing jobs and lead to serve unemployment.
 - Machines can easily lead to destruction if the implementation of machine put in the wrong hands the results are hazardous for human beings.
- [4] Robotic Process Automation
- Robotic process automation (RPA) is a software technology that makes it easy to build, deploy, and manage software robots that emulate human's actions interacting with digital systems and software. The technology is used for software tools that automate human tasks, which are manual, rule-based, or repetitive. Typically, it is like a bot that performs such tasks at a much higher speed than a human alone. These RPA software bots never sleep make zero mistakes and can interact with in-house applications, websites, user portals, etc. They can log into applications, enter data, open emails and attachments, calculate and complete tasks, and then log out. An RPA workforce is precise, accurate and immune to boredom. It can also be scaled more easily than your human workforce. RPA can perform just about any rule-based work and can do so through interaction with any software application or website. It's a robotic connection to the human world of the computer user interface. If a human can do it, a robot can do it in virtually the same way.
- [5] UiPath tool
- UiPath provides an RPA platform to automate digital business processes across front-end and back-end office tasks. It includes products such as studio, software robots, and orchestrator. It provides solutions in various sectors such as Banking, Finance, BPO (Business Process Outsourcing), Insurance, Retail, Telecom, Manufacturing, Healthcare, Public Sector, etc. It allows users to perform Web Automation, Desktop Automation, GUI (Graphical User Interface) Automation, SAP Automation, Mainframe Automation, Citrix Automation, Excel Automation, Screen Scraping, and Screen Recorder, etc.
- [6] Advantage of using RPA (UiPath)
- **Accessibility:** The extensible platform of UiPath provides hundreds of built-in, customizable, shareable activities and deep integrations with the help of various technologies that are already in use. The UiPath has mobile and browser accessibility.
 - **Rapid:** The ecosystem of the UiPath tool is optimized for faster development and designed in such a way to deliver the Quick return on investments.
 - **Artificial Intelligence:** The Artificial Intelligence robotic manager reduces the automation costs and meets the service level with the help of synchronized queue work and robot deployments with scheduled workflows and events.
 - **Scalability:** The RPA (Robotic Process Automation) at the enterprise level is expected to deploy and manage various variety and number of processes from front-end-office to back-end-office with regardless of complexity. The user can train tens, hundreds, or thousands of Robots at the same time by using the UiPath tool. This tool also has absolute consistency in job performance.
 - **Quality of the Agile process:** The Agile process is the technique that supports continuous iteration of development and tests the developing module throughout the whole software development life cycle of the project. The UiPath tool consists of an agile technique, which is very useful according to both client and organization.
 - **Flexibility:** Flexibility is the key advantage of the UiPath tool to build an effective digital workforce. The UiPath tool provides flexibility to both the user and the organizations.
- [7] Disadvantage of using RPA (UiPath)
- The UiPath tool of Robotic Process Automation improves the efficiency of organizations by reducing the repetitive human efforts, but there are some limitations of automated work. When automation is applied, then, it requires judgment related to work.
 - The RPA tool UiPath is not a cognitive computing solution.
 - This tool cannot read any data which is non-electronic with unstructured input.
 - When an enterprise uses the UiPath tool to automate the task, then the enterprise needs to be aware of several inputs that are coming by multiple sources.
 - The local hosting of the UiPath Orchestrator server is not available in the community edition of UiPath.
 - The main disadvantage in the UiPath tool is its auto-start feature of UIRobot.exe.
 - The Number of Robots is limited in the Orchestrator community edition.
 - The UiPath tool asks the user to activate the libraries from the nugget package manager, which gets deleted every time.

[8] What is invoice email scam?

Accounts payable (AP) departments at large enterprises pay thousands of invoices each month. It's difficult for AP teams to keep track of every invoice that comes in, making it tough to detect upward trends in pricing or the addition of surcharges over time. Suppliers can raise prices without warning or send multiple invoices for smaller amounts that are less likely to be scrutinized. With such a large volume of invoices, companies leave themselves at risk for overpayment and potential invoice fraud.

According to the "2020 Report to the Nations," published by the Association of Certified Fraud Examiners (ACFE), organizations lose an estimated 5% of annual revenue due to fraud. For organized scammers, accounts payable (AP) departments are often perceived as a poorly guarded cashbox they can target for scams or other malicious attacks. In most cases, payments are processed through these departments, which makes this function vulnerable and a prime target for theft.

[9] What are the frequent invoice frauds?

- **Fake Vendors and False Billing:** Most companies have long-term business contracts with vendors and suppliers. These are routine purchases; payments are often pre-approved without much inspection and scrutiny. Fake companies can mimic the actual vendors and send fake invoices. According to the ACFE report, fake billing is hard to detect and poses a significant risk, with a median loss of \$100,000. An employee can also engage in malicious activities by creating fake vendors and sending fake invoices. An example of this is an Oregon woman who embezzled millions of dollars from her company over nearly 15 years. Scammers can also work with internal employees to create payment requests for goods/services that the company didn't receive and then keep the processed payment for themselves. This is what happened with a former Honda employee in Ohio, who defrauded the company out of \$750,000.
- **Overbilling/Overpayment:** There are instances when a vendor deliberately introduces new line items to inflate the amount owed. When an AP process is manually driven and experiences a surge in invoice volume, these wrongly added line items can go unnoticed. And that results in the company getting overbilled.
- **Expense Account Malfeasance:** Expense account padding works a bit like fake billing. When employees submit expense reports for reimbursement on things like mileage, meals on the company dime, and office supplies, they're generally supposed to include receipts. But these documents aren't always closely checked and companies can easily wind-up rubber-stamping receipts that have no business getting approved. And around a quarter of businesses don't even require receipts for reimbursement.

[10] How to prevent invoice frauds using AI?

- **Automated Invoice Review:** Developing essential protection against accounts payable fraud involves matching key documents generated across the procurement stages. It includes matching invoice line-item information with purchase order details and reviewing receipts to ensure goods and services have been delivered. It often involves reviewing other communications with the vendor. When done manually, this can take days. But with AI-powered automation, the entire reconciliation cycle can happen in just minutes.
- **Employee Malintent Monitoring and Tracking:** Employees engaged with accounts payable fraud often operate as individuals or small groups, having gained strategic information about operations. Technology systems, like ML-based anomaly detection, can track and monitor employee engagements in the AP process in real time to thwart malicious behavior. Training employees on such procedures can increase awareness. Building a loyalty program can also help charter a course toward ethical actions.
- **Building Vendor Match Capability:** Since AP departments sometimes encounter invoices generated by fake companies, it's important to detect such activity at its initial stages and before damage is done. Detection can be done through fuzzy matching technology, in which the system scans multiple repositories, looks through the existing vendor database and checks various documents to detect fakes.
- **Automated Approval Workflow:** Automated approvals reduce the time-consuming manual approval processes. Automation of approvals ensures compliance by leaving a digital trail of the actions performed across each stage. Training and generating awareness among employees can help mitigate the risk of malicious intent.
- **Digital Solutions to Enforce Compliance:** Digitizing the entire invoice processing workflow can reduce AP fraud. Digitization calibrates actions with time stamps and personas, which can later be used to conduct system audits. Such systems can be designed to generate automated reports, where employee activities can be monitored and subsequent areas of improvement can be identified. Also, real-time alerts can give administrators holistic insights into process operations.

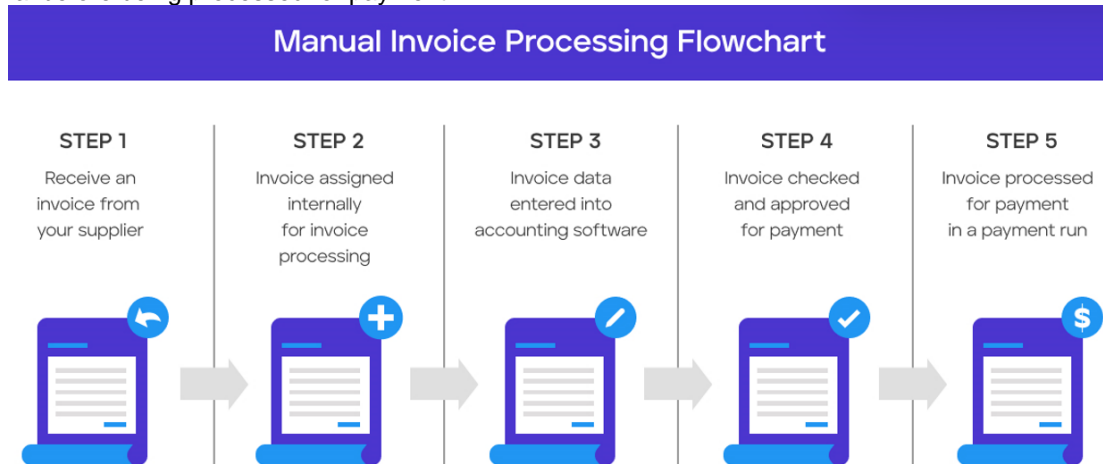
[11] What is invoice processing?

Invoice processing by definition is a business function performed by the accounts payable department which consists of a series of steps for managing vendor or supplier invoices from receipt to payment, and recorded in the general ledger. Invoice processing is often performed with software and it is commonly referred to as automated invoice processing or invoice automation for short. An invoice processing flowchart is a structured guide detailing the steps for how accounts payable is to process vendor invoices.

The following steps need to be performed for the invoice processing:

- **Step 1:** The traditional invoicing process starts with receiving paper invoices in the mail or PDF invoices via email or other electronic means from your supplier.
- **Step 2:** Next, that invoice is internally assigned for processing, which involves either scanning or manually entering the invoice data into your ERP system or accounting software.

- **Step 3:** From there, the invoice amounts must be checked and approved for payment, which can involve being coded for the right account, project, or cost center. If your business uses purchase orders, this may also involve PO matching.
- **Step 4:** Last but not least, the invoice is sent to the appropriate person, often a department head, for review and final approval before being processed for payment.



[12] Drawbacks of manual invoice processing

- **Affects Cash Flows:** Manual invoicing is time-consuming and boring, and that can lead to mistakes and delays in sending invoices to your customers at the right time which can affect getting paid quickly, thus affecting your cash flows. It is also difficult to track missed or delayed payments as you don't get to see the pending invoices at a glance.
- **It's costly:** On average, it costs about more to manually process just one invoice. If your organization manually processes hundreds and thousands of invoices annually, you can just do the calculations! Manual operation is not a good idea anymore.
- **Cannot use latest payment modes:** With manual invoicing you cannot use the latest mode of payments, ex. Online credit/debit card, UPI, QR Codes, PayPal, etc., which can significantly reduce payment receivable life cycles for your company and makes life easy for your customers too.
- **It's time-consuming:** With a manual system, it takes much longer to prepare and send an invoice. Your team spends more time to get less significant work done. This further contributes to higher cost of operation, as you'd have to pay more people for longer working hours.
- **It causes you to lose money:** Issues of missed invoices, delayed payments, wrong values and missed entries are all very common with manual invoice management. Whenever this happens, you lose money and your company's bottom line is impacted.
- **It's boring:** And often very annoying. This leads to ineffectiveness for your team. Nobody wants to sit at the same old desk and enter mind-numbing data for hours, every business day, but unfortunately, that's what manual invoicing is all about.
- **It's prone to human error:** Since processes are manned by error-prone humans in a manual system, mistakes are inevitable. They can even lead to bigger issues, which only means wasted time and money for your organization.
- **It's a lot of hassle:** Manual invoice processing is often complex and inconvenient. To prepare invoice one needs to extract information, match details, send it back and forth for different workflows like approvals, prepare checks, and do a lot of other complicated stuff.
- **It requires more hands:** A manual invoice processing procedure needs more human workers to function, given that the related processes are hand-operated. And you would have to pay all of those humans.
- **It delays progress:** In a manual invoice processing system, since your workers are always sitting on the same desk entering data or performing some other related tasks, their time is not freed up to work on more value-adding or growth-oriented action items. This slows down advancement and development, making it difficult to implement new strategies and make progress on other fronts.
- **You can't easily track your cash flow:** Without clear and quick information of pending invoice, you cannot ascertain a proper cash flow for your organization. A manual system doesn't let you track this very easily.
- **It's vulnerable to security risk:** A manual system is easily exposed to security problems like data theft, data loss, and the like. When stored locally, anybody can illegally access and use your data as they want. Data security is important because it represents the financial operation of your organization.

[13] Advantages of automated invoice processing

- **Time saving** – Reducing the amount of manual input required from the accounts payable team to be able to properly process incoming invoices means there's more time for them to spend on other activities, like building strong relationships with suppliers and working on further improving AP efficiency.
- **Cost saving** – Studies show that it costs up to \$11.57 to process a single invoice, on average. Streamlining the process and reducing manual requirements can reduce this cost, generating savings in the AP department. AP automation can

also generate savings in the form of making it easier to spot opportunities to access early payment discounts and reducing the chance of invoices going missing, therefore avoiding late payment fees.

- **Reduced errors** – The accuracy with which invoice processing is carried out is massively important. Errors caused by manual input can include overpayments and duplicate payments, both of which are harmful to the bottom line. Automating invoice processing can significantly minimize the chance of these errors occurring.
- **Fraud protection** – Payment's fraud is a constant risk with traditional methods of invoice processing, but automated invoice processing removes a lot of the danger. Access to certain abilities, like approving invoices or making payments, can be limited to just the right people through access controls, meaning fraud is significantly more complex to perform.
- **Easier auditing** – Finally, automating the invoice processing cycle means that there's a secure, backed-up trail of all AP activities that can be relied on for future audits and process reviews. Documents, invoices, receipts, and messages that are involved with the same transaction can generally be linked together, creating an easily trackable audit trail to follow.

[14] Automation of invoice processing using RPA (different tools and methodology)

- **Document understanding in UiPath:** In this methodology, we follow the **Optical Character Recognition (OCR)** technique and the **machine learning algorithm** in the back-end. This methodology has a number of steps that need to be followed. First, we create a taxonomy for each document type if required, after this pre-processing step we move first processing step i.e., the digitization step, in this step we digitize the document using the different OCR engines available. Next, we classify the document using some of the key phrases so that we don't accidentally process the wrong document. After classification, we move to the step of data extraction where we use ML algorithm for extracting data from invoices. After this step we have an option of validating the extracted output manually, this can be done using human validation station, this station helps us alter any data if required, this helps in achieving 100% accuracy. After all the data is validated, the output so extracted is stored in excel and can be used for further processing. The drawback of this methodology is that due to the addition of human validation station, it takes a little bit longer but this step helps guarantee that the extracted output is 100% accurate.
- **Screen Scraping in UiPath:** In this methodology, the pdf is opened via application (Acrobat Reader) and then anchors are used to recognize information. The drawback of this methodology is that, the automation might give error when unable to find the selector showcasing an anchor or any element. This methodology uses **Optical Character Recognition (OCR)** technique to find the information and elements.
- **Regex in UiPath:** In this methodology, the entire pdf is read via **Optical Character Recognition (OCR)** technique and is stored as a string, from the output string the necessary details are extracted via regex technique and then can be used for further processing by adding the extracted value to excel or any other file type as per requirement. The drawback of this methodology is that it is dependent on the regex expression, which might not be correct with different type of invoices.
- **IQ Bot in Automation Anywhere:** In this methodology, the invoice is processed using the tool called IQ Bot. IQ Bot is an intelligent document processing platform from Automation Anywhere that can be used to classify, extract, and validate content from documents. IQ Bot allows us to process unstructured data using AI technologies such as computer vision, **Natural Language Processing (NLP)**, **Machine Learning (ML)**, and text classification. IQ Bot converts unstructured data from invoices to structured data.

III. METHODOLOGY

The process diagram/methodology followed by the paper is as follows:

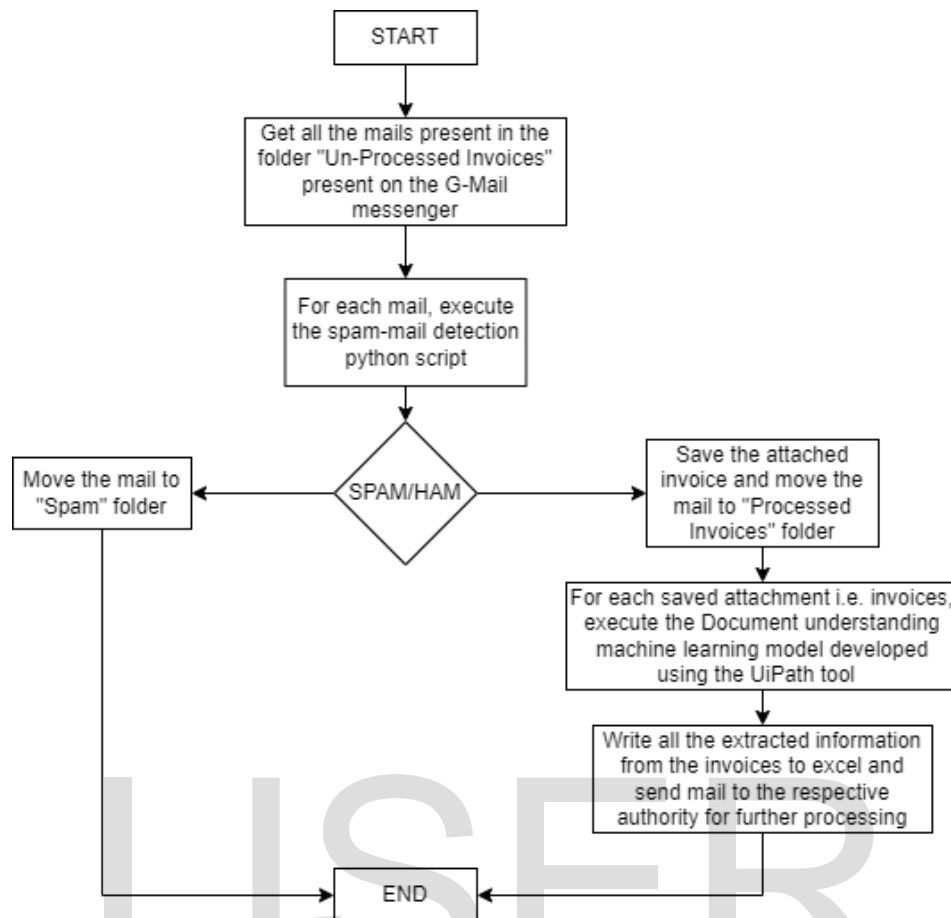


Figure 1 Process Diagram of the entire automation

In this automation, we have attempted to provide a secure and an accurate automation. The spam detector has an accuracy of 98.6% and the invoice data extraction part has the accuracy of more than 99.8%.

1) Spam Detection Process:

In this process we input the body of the mail and do analysis on the string and predict the type of the mail (spam/ham). The first step of this process is to create a model which helps predicting the status of the mail. We start by importing a number of libraries (Pandas, RE (regular expressions), nltk (count vectorizer, train test split, multinomial naïve bayes, confusion matrix, and accuracy score), sys (system), and sklearn (stopwords and porter stemmer). We then read the dataset that is used for training and testing the model, this is done using Pandas library. Next, we go to the step of data cleansing. In data cleansing first we remove all the characters other than lower- and upper-case alphabets, second, we convert all the alphabets to lower-case alphabets, third we split each sentence into words for easy removal of stopwords (stopwords are the words which do not help in the prediction), forth we find the root words for each word left and then we join each of the word into a string and append it to an array. Next, we go to the step of creating bag of words. We do that by initializing the countvectorizer with the max limit as 2500 and then fit-transform the new list. This step helps filter out the most frequently used words in the messages, this also helps in easy data cleansing. Next, we convert category column data to dummy data making ham as 1 and spam as 0. Next, we split the dataset into training and testing dataset. We have used naïve bayes classifier as it works really good with NLP (natural language processing). Next, we fit multinomial naïve bayes with X and Y training dataset in order to create the model. We can use confusion matrix and accuracy score to check the no. of values correctly predicted and also to get the accuracy percentage. After the model is developed, the main task starts. We get the input from user via command line argument (the input is the mail body that needs to be processed), after getting the input we perform the data cleansing steps and also the bag of words step. After that we apply model to the list and get the prediction stored in a text file for further use.

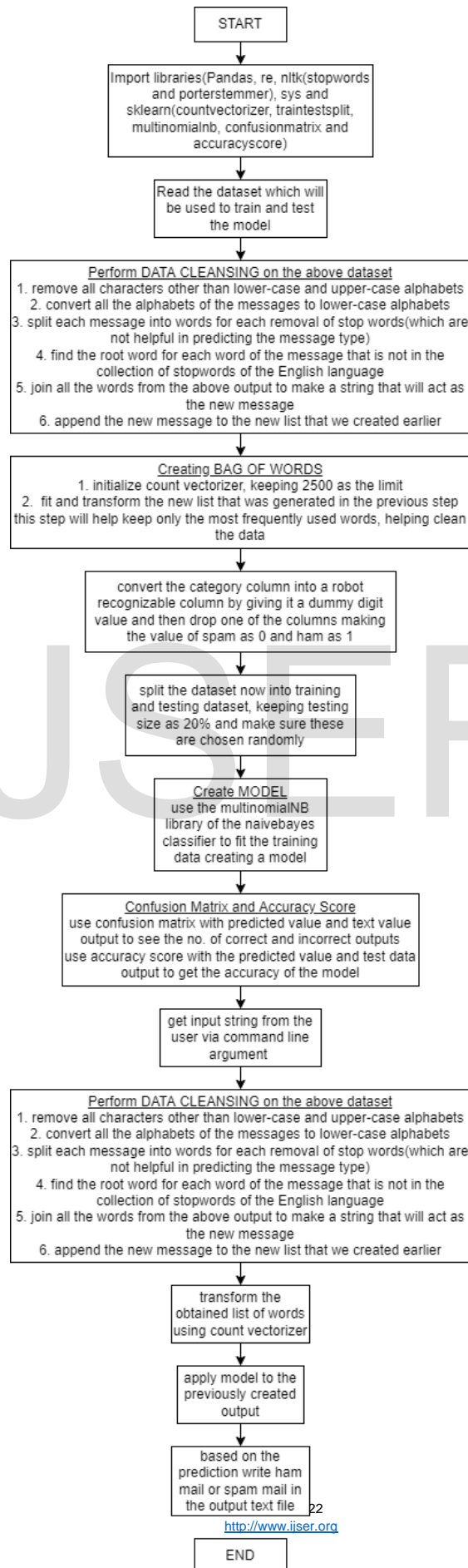


Figure 2 Process Diagram of Spam-mail detection process

2) Invoice Data Extraction Process:

In this process we extract all the relevant data there is from the invoice and consolidate it in an excel for the particular date. The first step is to create a taxonomy, specified according to the data that needs to be extracted. Next, we load the taxonomy and digitize the document using OCR engine. We then classify the document, so that for the further steps only invoices are sent. Post this step we perform data extraction using ML extractors, they have built-in machine learning algorithms. This step extracts all the data from the invoices which can be cross-checked in the next step of validation station, where the person can check and validate each extracted value, so that no value is incorrect in the final output. The approved data is then exported into datasets which then can be written in excels and mailed to the respective user.

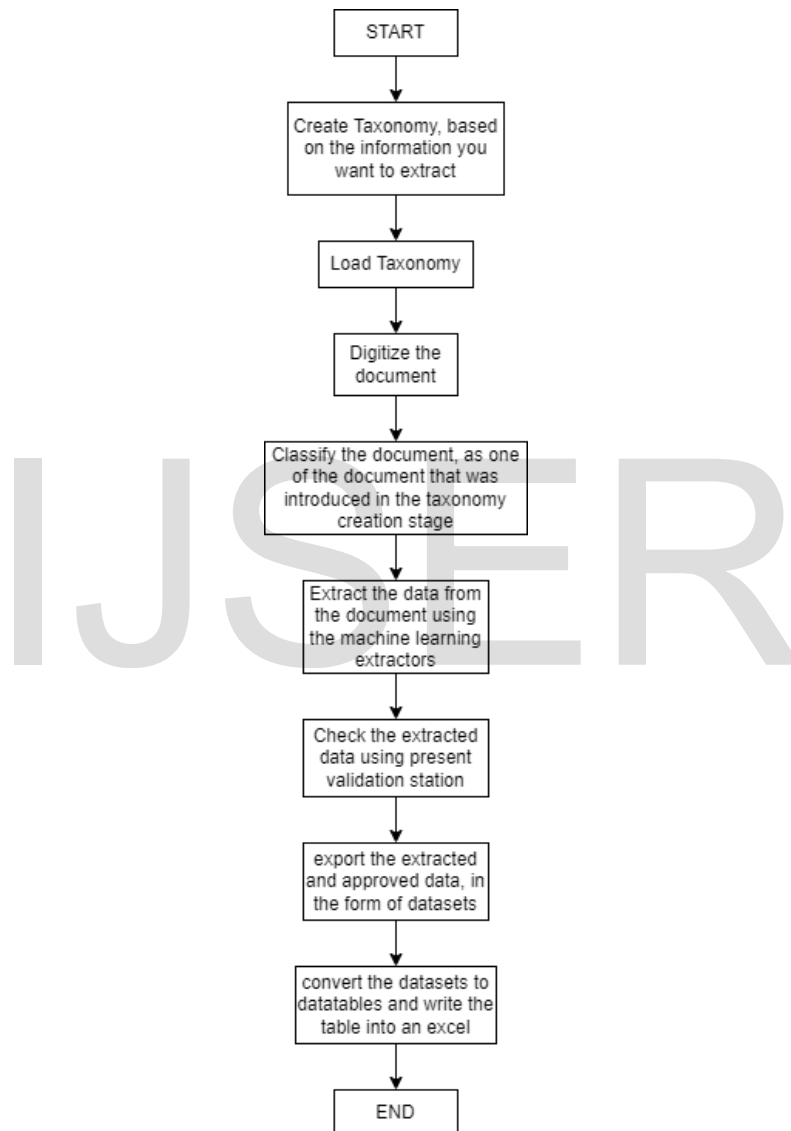


Figure 3 Process Diagram for Invoice Data extraction process

IV. LITERATURE REVIEW

The invoicing process is an important part of a wider set of business processes including the placing and acceptance of an order, delivery and payment. Some research papers regarding invoice processing system were studied by us. These included various techniques of invoice processing using cognitive approach, generic system which uses OCR engine, using CBR (Case-based reasoning), optimization approach which makes use of SVM, Entropy, using IQ Bot of Automation Anywhere, using regex technique of UiPath, etc. Some of these approaches for invoice processing system are mentioned below.

Sagar Sahu, Sania Salwekar, Atharva Pandit, Manoj Patil in their paper proposed the methodology of using UiPath's inbuilt OCR engines for example Microsoft OCR, UiPath OCR, Tesseract OCR, etc. They extracted the data from the invoices using the

OCR engine and then used the regex expressions to extract relevant information from the string and fed it to excel for further processing.

Devanshi Desai, Ansh Jain, Dhaivat Naik, Nishita Panchal, Dattatray Sawant in their paper proposed the methodology of using Automation Anywhere's IQ Bot. IQ bot is used to extract structured or unstructured data as it uses AI, ML, NLP to learn and extract information from the documents. The reason that they gave for choosing this is that it has AI-based learning algorithms to recognize and classify the content and the fact that it doesn't depend on OCR is a really good plus point. Also, because it's an IQ bot, it continuously learns from user feedback and validation therefore increases accuracy. They used ML based IQ Bot. The IQ bot contains a ML tool using Naïve Bayes Algorithm. Bayesian Naïve Bayes algorithm is based on Bayesian decision theory and is widely used. The algorithm is built on features that it employs to identify the text in question.

Enes Aslan, Ethem Unver, Tugrul Karakaya, Yusuf Sinan Akgul in their paper proposed a new invoice parsing method which consists of two-phase optimization structure and eliminates invoice classes. The first phase uses individual invoice part detectors such as SVM, maximum entropy and HOG to produce candidates for the various parts of different types of invoices. At the second phase, the basic idea is to divide an invoice into different parts and then arrange it together. As PBM is an optimization-based method, it can handle any type of invoices. The proposed system is tested with real invoices and are found to be promising for the real-world experiences.

V. REQUIREMENT ANALYSIS

Requirement's analysis involves all the tasks that are conducted to identify the needs of different stakeholders. Therefore, requirements analysis means to analyze, document, validate and manage software or system requirements. High-quality requirements are documented, actionable, measurable, testable, traceable, helps to identify business opportunities, and are defined to a facilitate system design. After the extensive analysis of the problems in the system, we are familiar with the requirement that the current system needs. The system requirements are categorized into the functional and non-functional requirements. These requirements are listed below:

- A. Hardware requirements
 - a. Minimum 4GB RAM.
 - b. 200 MB of free Hard Disk space.
 - c. Browser: Chrome(v49), Internet Explorer (v10) or higher.
 - d. Processor: 3GHz or higher

- B. Software requirements
 - a. Operating System: Windows 7 and above.
 - b. UiPath Studio.
 - c. UiPath Robots.
 - d. UiPath Orchestrator.
 - e. Python.

- C. Functional requirements

Functional requirement are the functions or features that must be included in any system to satisfy the business needs and must be acceptable to the users. Based on this, the functional requirements that the system must require are as follows:

- a. **Customer Name:** The Name of customer is required to identify the person who has initiated the transaction. The name should be spelled correctly in order to avoid confusion.
- b. **Total Amount:** This field gives us the final amount of the product.
- c. **Invoice Number:** This field shows us the Receipt (Invoice) number of the transaction. This number is useful in servicing centers to identify the product warranty or guarantee period.
- d. **Invoice Date:** Date at which the invoice was printed.
- e. **Product Description:** This field gives us the information about either products or services including prices and quantities. Often includes standard product description and inventory number.
- f. **Currency:** Currency of the total amount to avoid any confusion during the time of payment or calculations.

- D. Non-functional requirements

Non-functional requirement is a description of features, characteristics and attribute of the system as well as any constraints that may limit the boundaries of the proposed system. The non-functional requirements are essentially based on the performance, information, economy, control and security efficiency and services. Based on these the non-functional requirements are as follows:

- a. **Security:** - Security requirements ensure that the software is protected from unauthorized access to the system and its stored data. It considers different levels of authorization and authentication across different user's roles. For instance, data privacy is a security characteristic that describes who can create, see, copy, change, or delete information. Security also includes protection against viruses and malware attacks.
- b. **Reliability:** - Reliability defines how likely it is for the software to work without failure for a given period of time. Reliability decreases because of bugs in the code, hardware failures, or problems with other system components. To measure software reliability, you can count the percentage of operations that are completed correctly or track the average period of time the system runs before failing.

- c. **Performance:** - Performance is a quality attribute that describes the responsiveness of the system to various user interactions with it. Poor performance leads to negative user experience. It also jeopardizes system safety when it's overloaded.
- d. **Availability:** - Availability is gauged by the period of time that the system's functionality and services are available for use with all operations. So, scheduled maintenance periods directly influence this parameter. And it's important to define how the impact of maintenance can be minimized. When writing the availability requirements, the team has to define the most critical components of the system that must be available at all time.
- e. **Scalability:** - Scalability requirements describe how the system must grow without negative influence on its performance. This means serving more users, processing more data, and doing more transactions. Scalability has both hardware and software implications. For instance, you can increase scalability by adding memory, servers, or disk space. On the other hand, you can compress data, use optimizing algorithms, etc.

VI. TECHNOLOGIES USED

A. UiPath Studio

UiPath Studio is a complete solution for application integration, and automating third-party applications, administrative IT tasks and business IT processes. One of the most important notions in Studio is the automation application.

An application is a graphical representation of a business process. It enables you to automate rule-based processes, by giving you full control of the execution order and the relationship between a custom set of steps, also known as activities in UiPath Studio. Each activity consists of a small action, such as clicking a button, reading a file or writing to a log panel.

The main types of supported workflows are:

- a. Sequences - suitable to linear processes, enabling you to smoothly go from one activity to another, without cluttering your workflow.
- b. Flowcharts - suitable to a more complex business logic, enabling you to integrate decisions and connect activities in a more diverse manner, through multiple branching logic operators.
- c. State Machines - suitable for very large workflows; they use a finite number of states in their execution which are triggered by a condition (transition) or activity.
- d. Global Exception Handler - suitable for determining the workflow behavior when encountering an execution error, and for debugging processes.

B. E-Mail Activities Package

- a. The Mail Activities Pack is designed to facilitate the automation of any mail-related tasks, covering various protocols, such as IMAP, POP3 or SMTP. UiPath also features activities that are specialized for working with Outlook and Exchange.
- b. Activities such as Save Mail Message and Save Attachments are not intended to be used with certain mail protocols. Instead, they save the MailMessage object variable retrieved from activities such as Get IMAP Mail Message to a specified folder on the current machine.

C. Document Understanding Package

The activity package that enables Machine Learning Document Understanding features, such as the Machine Learning Extractor and the Machine Learning Classifier. The UiPath Document Understanding Framework is designed to help users combine different approaches to extract information from multiple documents, not necessarily with the same structure.

D. Excel Activities Package

- a. The Excel activities package aids users to automate all aspects of Microsoft Excel, as we know it is an application intensely used by many in all types of businesses.
- b. It contains activities that enable you to read information from a cell, columns, rows or ranges, write to other spreadsheets or workbooks, execute macros, and even extract formulas. You can also sort data, color code it or append additional information.

E. UiPath Robot

- a. The Robot is an execution agent, meaning that you have to provide it with the automation applications you want it to run.
- b. After creating an automation application in Studio, it needs to be published locally or to Orchestrator. Once an application is published, you can send it to the Robot machine and start executing it.
- c. This is populated by default as follows:
 - i. When NOT connected to Orchestrator
 - ii. When connected to Orchestrator - the default Orchestrator feed

F. Intelligent OCR Activities Package

The package contains core activities that enable the usage of a complete document processing framework, from taxonomy definition, digitization, document classification, data extraction, data validation and classifier / extractor training. Taxonomy can also be created specifically for each type of document.

G. Python software

The Python programming language is a popular programming language because of its simplicity, ease of use, open-source licensing, accessibility, renowned community, great support and help, tons of packages, tutorials, and sample programs – which make it easy for a beginner to learn and code with Python. Python can be used to develop a wide variety of

- 2) The invoices are received via email in the particular mail folder.

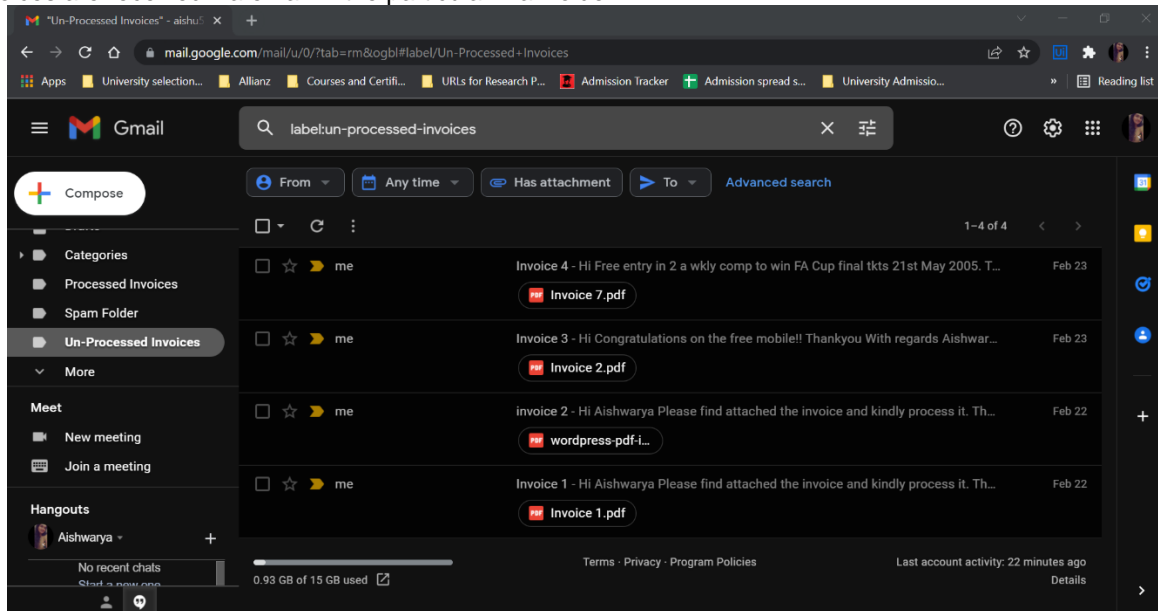


Figure 5 Invoice mails in the mailbox

- 3) All Invoices sent to UiPath Robot through email is successfully downloaded to a dedicated folder which is assigned during development. The robot never redownload the invoice which has been already downloaded previously. These invoices were downloaded only after they got the green flag from the spam detection python script with 98.6% accuracy.

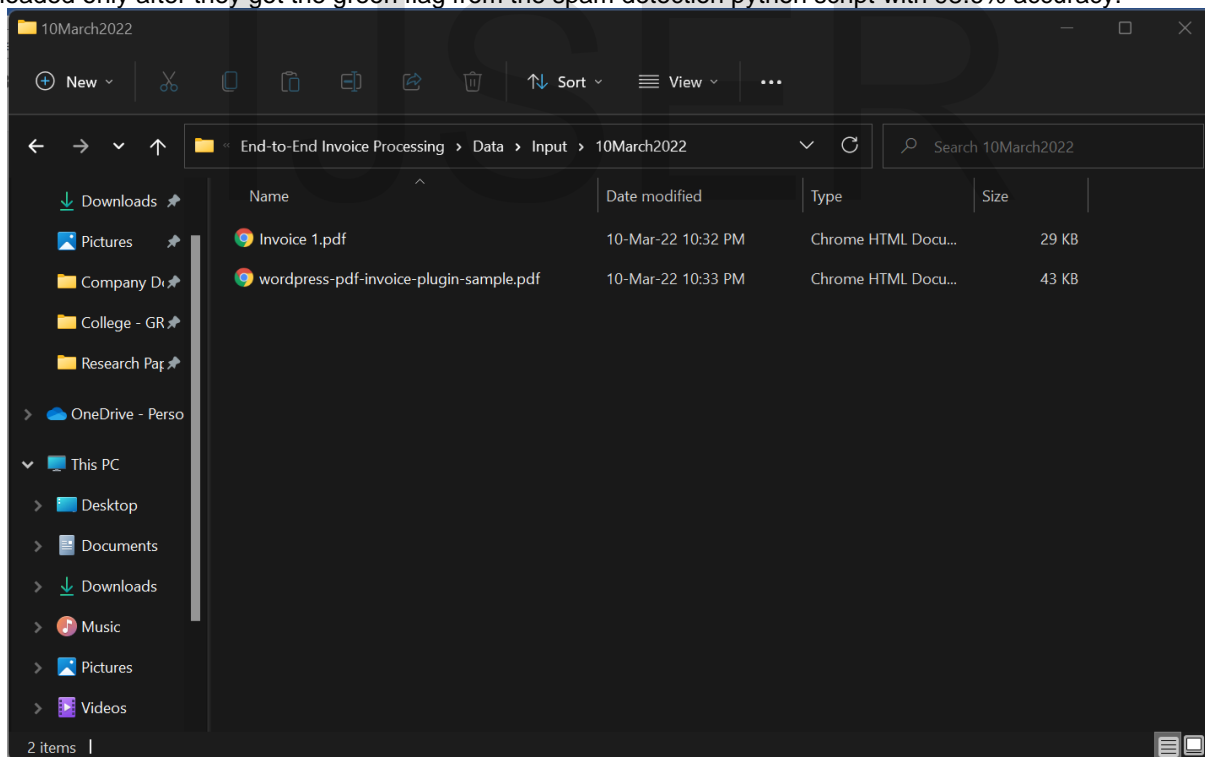


Figure 6 Successful download and storing of invoices in the current date folder

- 4) After successfully downloading the invoices, the robot correctly fetches invoices from that folder one by one. After this the robot digitizes the document for OCR engine to work accurately and then checks if the document is invoice or not and if it is an invoice, it extracts data from the pdf and stores it in an excel with 100% accuracy.

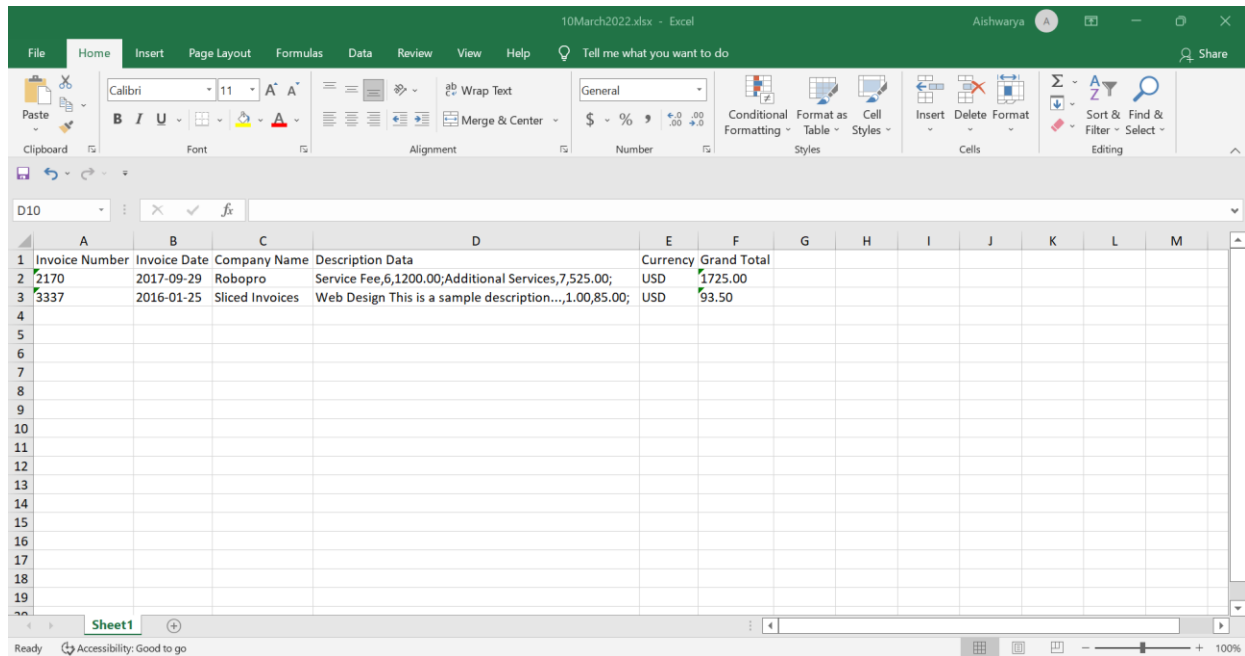


Figure 7 : Successful data entry of content in the Excel file

- 5) After successfully registering each invoice, the software robot is then able to send post notifications in the form of email to the concerned employee or to the vendor in question.

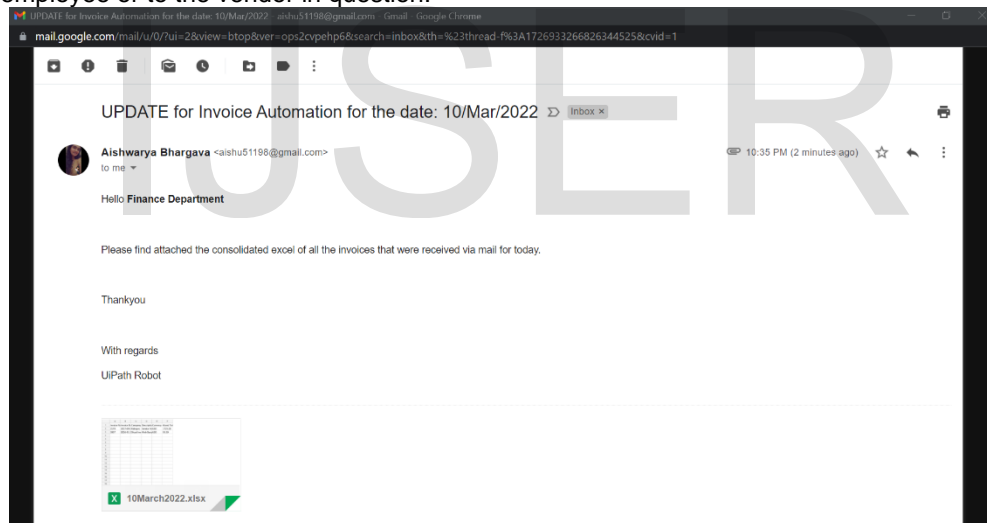


Figure 8 Successfully notifying the concerned employee or vendor

- 6) The spam mails are transferred to the spam folder of the mail box.

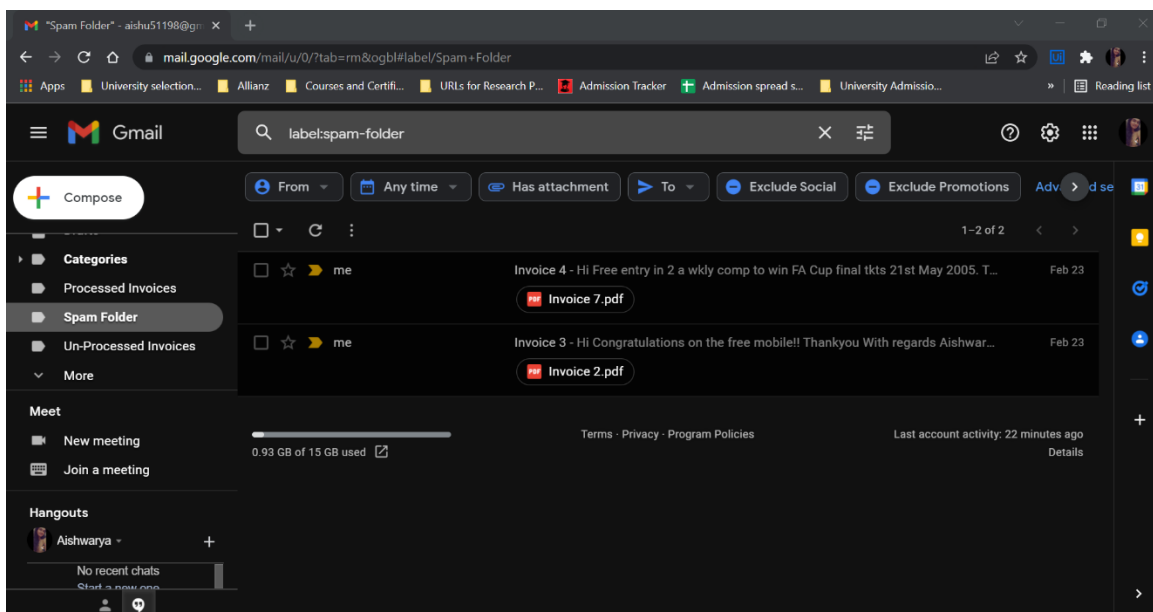


Figure 9 Spam folder with the spam mail-invoices

7) The processed invoices are transferred to the processed invoices folder of the mail box.

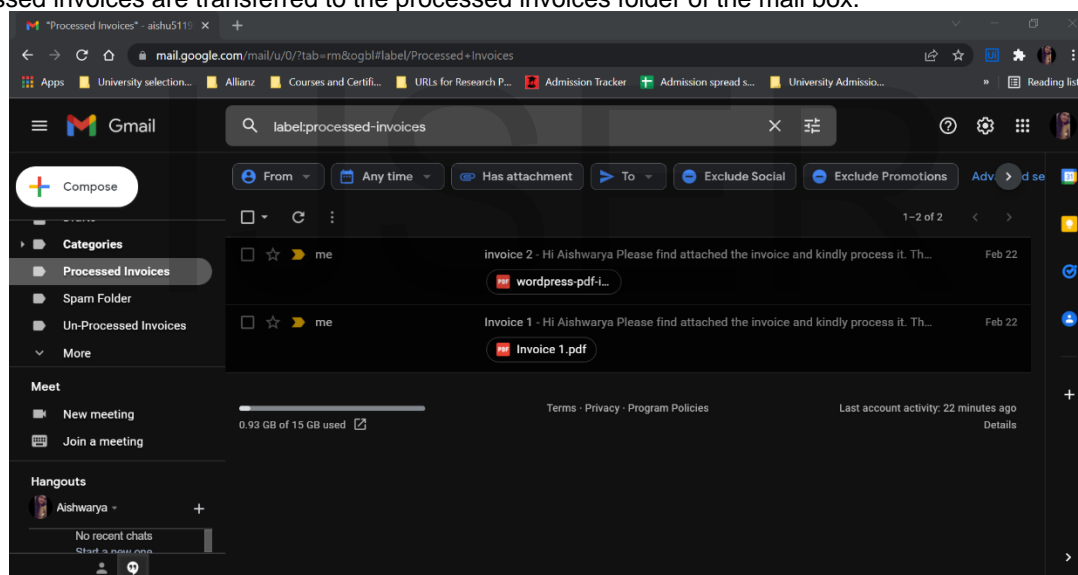


Figure 10 Processed Invoices folder with the mails and invoices that have been processed by the robot

VIII. SWOT ANALYSIS

Strength, Weakness, Opportunities and Threats, SWOT, is a framework used widely to assess the businesses four pillars. These four pillars help understand about all there is to be known. This 2x2 matrix, created after brainstorming every block in relation to our product, will tell us where we stand.

<p style="text-align: center;"><u>STRENGTHS</u></p> <ul style="list-style-type: none"> • The automated invoice processing is done using machine learning and document understanding tool present within UiPath Studio software. • Machine learning helps us remove all the spam invoices from the list, hence helping us extract information from ham mails/invoices only. • Document understanding tool allows us to extract information from various documents using machine learning and different intelligent technologies as well. 	<p style="text-align: center;"><u>WEAKNESSES</u></p> <ul style="list-style-type: none"> • Even though it is automated, it will require human intervention in order to achieve 100% accuracy for data extraction. • UiPath Studio needs to be installed on the system in order to execute the automation along with all the python libraries and python software. • This could make them slow and could delay the further dependent processes.
<p style="text-align: center;"><u>OPPORTUNITIES</u></p> <ul style="list-style-type: none"> • More than 80% to 90% of the data in any organization are placed within electronic documents and forms. • So, processing documents is one of the most common requirements for most automation use-cases and also this processing needs to be only done for ham documents and with a lot of security. • The company has the opportunity to expand its global operations and can focus on letting their employees work on much more important and critical operations. 	<p style="text-align: center;"><u>THREATS</u></p> <ul style="list-style-type: none"> • Who knows if the need for Automated Invoice Processing System will grow and stay in favor with companies, or whether another type of Invoice Processing system is introduced that will replace the Automated Invoice Processing System? • There is a high risk of missing out on processing the invoice if it isn't present in the specified mail folder. • Its success has led to the market entry of many competitor companies and copycat brands that pose potential threats.

Figure 11 SWOT Analysis Diagram

IX. CONCLUSION

Automated invoice processing can achieve powerful results for accounts payable departments. Thanks to technological advancements in robotic process automation and computer vision technologies, invoice processing can eliminate bottlenecks within the AP process and turn the department into the profit center it can be. Automated invoice processing enables touchless automation across the entire accounts payable process, and can transform the business in just months, creating a powerful return on investment. Any organization that receives a large number of vendor invoices on paper can benefit from invoice processing technology. The more data from each invoice that you are hand-keying into your accounting software that more benefits you can get from each page you automate.

With the help of the solution discussed above the finance department of any company would be effectively able to prevent from processing any spam invoices and will also be able to save a lot of time as they will not have to do anything manually. The solution takes care of filtering the spam/ham mails, and also of extracting the required data and also sending the extracted results excel to the concerned person. This solution is secure, effective, efficient and accurate.

X. BIBLIOGRAPHY

- [1] Sagar Sahu, Sania Salwekar, Atharva Pandit, Manoj Patil, "Invoice Processing Using Robotic Process Automation", 2020, International Journal of Scientific Research in Computer Science, Engineering and Informational Technology.
- [2] Desai, Devanshi and Jain, Ansh and Naik, Dhaivat and Panchal, Nishita and Sawant, Dattatray, "Invoice Processing using RPA & AI", 2021, International Conference on Smart Data Intelligence
- [3] Harikrishnan NB, Vinayakumar R, Soman KP, "A Machine Learning approach towards Phishing Email Detection", 2018, ACM IWSPA
- [4] Jorge Ribeiro, Rui Lima, Tiago Eckhardt, Sara Paiva, "Robotic Process Automation and Artificial Intelligence in Industry 4.0 – A Literature review", 2020, CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies
- [5] Ravi Teja Yarlagadda, "The RPA and AI Automation", 2018, International Journal of Creative Research Thoughts
- [6] A. SARAVANAN and S. SATHYA BAMA, "A Review on Cyber Security and the Fifth Generation Cyberattacks", 2019, Oriental Journal of Computer Science and Technology