Implementation of a Web System: Prevent Fraud Cases in Electronic Transactions

Edwin Kcomt Ponce, Katherine Escobedo Sanchez, Laberiano Andrade-Arenas
Facultad de Ciencias e Ingeniería
Universidad de Ciencias y Humanidades
Lima, Perú

***Abstract*—The purpose of this project is to prevent cases of fraud in e-commerce of purchase and sale from person to person through social networks. For the development of the research work, the Scrum methodology was used to allow the project to be carried out in an agile and flexible way, adapting to the changes that could arise along the way. The technological tools that made this project possible were SQL Server, C++, Visual Studio and Marvel app, the latter for prototype design. In addition, there was the support of an artificial intelligence software known as Optical Character Recognition that allowed the document recognition process to be completed. The social network Facebook was also relevant for the development process since the data set for the training of the system was obtained from there, guaranteeing its functionality. The results obtained benefit both parties, sellers/suppliers and consumers, reducing the impact of fraud cases and guaranteeing safer online operations. In addition, a validation was carried out by experts in the development of web applications, taking usability, feasibility, scalability, innovation, and technology as criteria. Obtaining as a result the approval in all its criteria; with the total mean value of 2.76.**

***Keywords*—Artificial intelligence; e-commerce; fraud; optical character recognition; scrum; social networks; web system**

I. Introduction

Digital transformation has become more relevant in recent years as a result of the pandemic, although this process brings with it greater benefits and a positive impact on the development of a country, it also poses certain risks for a large number of the population. According to the World Economic Forum in a survey presented in its Global Risks Report 2021 [1], it indicates that cybersecurity flaws rank 7th out of a total of 37 risks presented. In Latin America, the growth of e-commerce due to the pandemic had an exponential increase, which meant a greater use of devices with internet access to carry out various operations such as payments or purchases of goods and services online [2]. Approximately 10 million Latinos have purchased consumer goods through electronic stores, however, one of the most relevant problems is the criminal activities carried out through computerized means as a consequence of this new normal [3].

Peru also adds to this growth, with e-commerce operations reaching an increase of 86% in June 2020 and reaching its highest point in July with 160%. This represents approximately 6 billion dollars, which covers the size of the e-commerce market as detailed in the report by the Peruvian Chamber of Electronic Commerce. However, this led to an increase in claims related to online transactions due to the few regulations and measures in this area. Let's take into account that before the pandemic, Peru had around 65,700 businesses that sold online, but by the end of 2020 there were 263,200 businesses. We must add that this also brought with it the increase in informality in the digital sector [4].

In Peru there is a high rate of informality and this has been increasing according to the latest data provided by the National Institute of Statistics and Informatics, reaching an informality rate of approximately 75%, also affecting electronic commerce [5]. It is mainly social networks that are used for this type of informal trade, where many independent workers offer their products and services at very attractive prices. In this way they manage to capture even more the attention of the public, this activity through social networks represents an opportunity to generate income without discounts or commissions when carrying out their transactions. However, it is also an opportunity for fraudsters to carry out their criminal activities hiding behind anonymity and lack of security [6] [7]. These types of crimes are carried out through online transactions for advance payments without the guarantees or security that a platform of a formal company offers. The purchase and sale operation through social networks is based solely on trust between the seller/supplier and the customer, thus exposing themselves to being a victim of fraud [8].

Given the foregoing, the aim is to offer users the alternative of carrying out their online purchase and sale operations with greater security, thus avoiding and reducing computer fraud rates. This not only benefits the buyer but also the seller as it is an opportunity to demonstrate that their services and/or products can be requested with complete confidence. Thus, it also contributes to the security of the assets and personal assets of the citizen covered by the Law on Computer Crimes, thus being a reference to avoid fraud either by advertisements on a website or by social networks.

The objective of this research work is to implement a web system to prevent fraud in online transactions carried out by people through social networks. The platform will be available free of charge, allowing users to register and contribute by reporting fraud cases in order to expand the database, which in turn favors the system by improving its efficiency and accuracy. So that the reports by complaint are valid and avoid cases of misinformation; The user will be asked to attach the respective complaint so that their report can be considered. In this way, the aim is for users to have a system that helps them guarantee greater security in their payment transactions or online purchases for a requested product. As well as independent workers or business owners, they can avoid loss of resources and time by ensuring that their sale

is finalized and that it is not canceled at the last minute.

The article is organized as follows: in Section II the review of the literature, Section III the methodology, Section IV the case study, Section V results and discussions, Section VI the conclusions and finally Section VII future work.

II. Literature Review

The investigation addressed the issue of fraud prevention through electronic commerce, through the development of a web system. The studies of different authors on the research topic are analyzed in order to have information on trends, limitations, among others.

According to the authors [9], their application allows identifying customers who may pose a risk of fraud for an electronic commerce, through the information of cookies or IP configured in the devices used for online transactions. Normally, fraudulent clients do not make changes to their laptop, smartphone or computer, so the application manages to keep a record and identify the devices with a history linked to these criminal acts. To achieve their objective, the developers of the project use three main components for their fraud detection system, which are PC identifier, Address identifier and Asset Classifier. These methods are based on data extraction and statistical analysis. A total of 8,020 purchase requests were analyzed, efficiently detecting suspicious transactions compared to traditional methods that use data mining, but with poor performance.

For the [10] authors, the most frequent fraud operations are carried out through the use of a credit card through e-commerce. His proposal consists of applying the Support Vector Machine (SVM) algorithm, allowing fraudulent operations to be identified and classified from safe operations. The methodology for its system includes using fingerprint scanning as a measure to reinforce security during registration and access to online operations. If a transaction is classified as fraud, the payment process will be canceled and the fraudulent user's information will be sent to the database. Finally, it can be concluded that the SVM is an efficient algorithm with an accuracy of 99.9% according to a comparison table of several models prepared by the authors. These qualities of the algorithm make an accurate classification in the process of online fraud unlike other methods.

In the work developed by [11], the authors propose a new data intelligence technique with the aim of maximizing the model in the detection of fraudulent operations, without having to depend on the data variation that may exist. As in the previous article, this project focuses on operations carried out through online payments through e-commerce. The proposed system works with the Multiple Prudential Consensus model that uses and integrates the efficiency of different classification algorithms through a double criterion, probabilistic and majority. The final algorithm is determined to classify authorized or fraudulent operations according to the previously determined model and according to the analysis of criteria elaborated. The results obtained demonstrate the effectiveness of the model, compared to other systems used to detect cases of fraud. Of a total of 492 fraud samples, the developed model was able to correctly detect 394 compared to the 349 cases that were detected by the model that ranked second. The results confirm that the proposal stands out among many other solutions both in terms of models and classification.

The authors [12] use Blockchain-based technology which is quite secure and efficient. One of its great advantages is that once the data has been registered, it is possible to alter or delete it. In addition, it is a fairly simple technology and easy to understand. Blockchain is a chain of blocks which consists of a single registry network where information from the previous block is stored and thus passes the information to the following blocks. Based on this they decide to implement this technology to support online fraud detection and thus qualify a valid purchase. Blockchain will store verified transactions and their associated qualifications, making "verified" labels no longer necessary. It is concluded that blockchain systems are quite efficient in detecting online user fraud compared to traditional methods. In addition, the system allows better control of false accounts or information and has a reputation system to effectively reduce the number of fraudulent ratings.

In the research work developed by the authors [13], they propose a proposal based on Machine Learning in order to improve the existing fraud detection systems. They consider that there is an urgency for a better detection and prevention of cases of electronic fraud. Its model is based on the use of Machine Learning algorithms integrating big data, allowing it to predict the probability that an electronic operation is safe or fraudulent. The model was trained with a dataset composed of history of credit card transactions through electronic commerce; in order to predict any probability of fraudulent operations. Specifically, supervised learning algorithms such as Random Forest, Support Vector Machine, Gradient Boost and combinations of these were used, comparing their performance. The results obtained confirm the accuracy and precision of the proposal by combining the Gradient Boost + Logistic Regression algorithms. In addition, the model, being based on Machine Learning, has an addition that is Active Learning. This new addition allows solving data labeling problems, which means an improvement in learning for fraud detection.

After analyzing the different proposals developed by the cited authors, it is considered that the use of technologies and application of engineering is of the utmost importance to combat the growing increase in cases of computer fraud. Although most of the works focus on the use of algorithms and models for more sophisticated systems, none addresses the problem of fraud when both parties involved are natural persons. The approaches and initiatives apply to companies already established and that have the economic factor to implement the systems in their respective businesses. However, in our country there is a large percentage of independent workers who offer products through social networks. Merchants do not have their own payment system and are also responsible for making deliveries in person via delivery. For this reason, what we are looking for is to develop a web system that facilitates the safe use of this type of electronic commerce. Thus, users can make their purchases with the security of knowing if they are dealing with someone reliable for their transactions.

III. Methodology

In the research carried out, an agile methodology was used, which allows the development team to have constant

communication and be able to carry out work efficiently, quickly adapting to the changes that may occur during the project. Next, the methodology used and the technological tools that were necessary for this project are described in detail.

A. SCRUM Methodology

It is a methodology that offers a personalized and flexible way of working that is suitable for software development projects with a variety of requirements [14]. This iterative work model establishes the delivery of project progress incrementally, thus achieving more efficient results and greater productivity. In Fig. 1, we graphically observe the process followed by this methodology and its respective work phases.

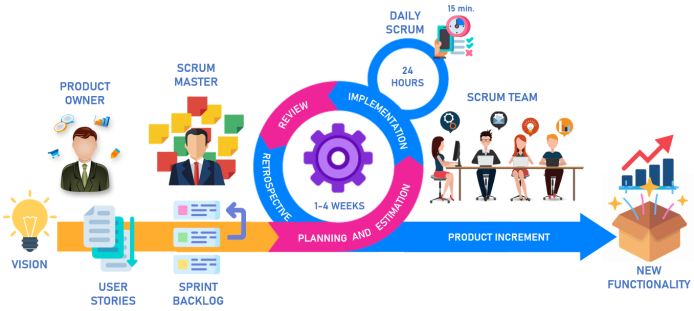

Fig. 1. Scrum Methodology Workflow.

1) Beginning: It is the first phase where the roles and functions of each member of the development team are identified and defined, these are assigned according to the skills and contributions of each member to the project, in the process the Scrum Master, Product Owner and the Scrum Team intervene [15]. Where the Scrum Master is mainly responsible for removing obstacles to the development of the project while the Scrum team is made up of developers, evaluators and other professionals necessary to guarantee the quality of the product.

2) Planning and Estimation: It is the next phase, the user stories are proposed and chosen according to the client's requirements, the product backlog is also carried out taking into account an estimation process for an adequate order of the stories. Once the requirements are organized, the sprint backlog is created. These are selected by priority and will go through the development and execution process. With the defined sprints, the team is responsible for between 1 and 4 weeks of the process to program, design and execute the sprint after the product increment is finished, the team continues with the next sprint.

3) Implementation: It is the third phase of the Scrum workflow, the objective is to deliver each sprint of the product organized, error-free and potentially operational, and generally at the beginning of the day during this phase small meetings are held [16]. These meetings are known as Daily Scrum that can last 15 minutes where the progress of the previous day is communicated and at the same time what impediments may be occurring for the progress of the project are discussed.

4) Review and Retrospective: It is the fourth and final phase of the work method, once the sprint is potentially deliverable at the end, a review of said sprint is carried out with the Product Owner to show the increase in the product. The increment is inspected and its functionalities are demonstrated. Sometimes the product list must be adapted according to the possible new requirements that the Product Owner indicates [17]. After the review, we proceed with the retrospective; where the Scrum Team makes an analysis of itself with the possibility of proposing strategies to execute improvements in the way of working for the following advances.

B. Technological Tools

For the research work, it was considered to use specialized applications and programs for this type of project with the intention of having an adequate development environment; thus achieving the correct implementation of the proposed system.

1) SQL Server: It is a relational database management system that uses the Transact-SQL development language. It is ideal because it allows you to store all the information you want with a wide variety of processes and with different utilities. Plus, easily integrate application data and leverage a rich set of cognitive services to power AI processes at any data scale.

2) Visual Studio: It is a program that provides us with an integrated development environment which facilitates the creation, design and development of web sites and applications, at the same time allowing us to work in environments that support the .NET framework. It is compatible with a wide variety of programming languages, such as C#, C++, Visual Basic, Python, Java, PHP among others.

3) Python: It is a high-level programming language that processes all kinds of data. Its software is free, that is, it has no cost, allowing it to be used and distributed even for commercial use. It is accessible and multiplatform, it has an extensive library, as well as a varied repertoire of frameworks, also standing out for its simplicity of syntax.

4) Marvel App: It is a web application to work online that allows us to make layouts and prototypes of both web pages and applications on mobile devices. The tools it makes available are sufficient to create designs that allow developers to have clear ideas of the final product.

5) Artificial Intelligence (AI): This technology aims to allow software to have the ability to learn based on data, which arise from patterns and opportunities provided by developers. Once the necessary data set is obtained, performance tests are carried out to measure and calculate the efficiency and accuracy of the AI software based on the number and quantity of hits and misses [18]. Deep Learning is a type of AI that uses multiple processing steps, also known as layers, to learn and then recognize data representations with multiple levels of abstraction. It is based on artificial neural networks, one of the specialties of this AI is image processing.

In Fig. 2, it can be seen how AI is classified, while in Fig. 3, we see the main classes of neural networks used by deep learning. For the development of this project, we worked with the Recurrent Neural Networks architecture because it was the one that best adapted to the required functionalities of the web system.

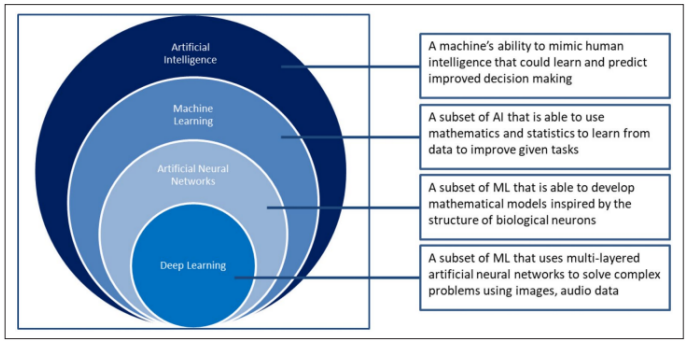

Fig. 2. Graphic that Shows how Artificial Intelligence is Classified in its Different Subfields [19].

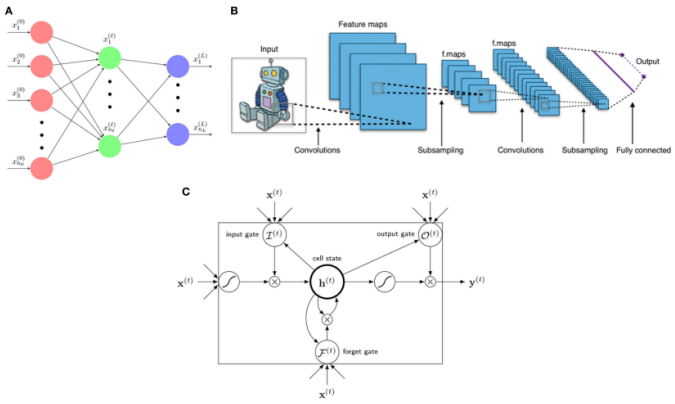

Fig. 3. The 3 Main Architectures of Deep Learning, A) Artificial Neural Network, B) Convolutional Neural Networks and C) Recurrent Neural Networks [20].

IV. Case Study

A. Planning

With the roles already defined and assigned among the team members, the planning continues. Where we choose the user stories, these determine the desired features and functions of the web system requirements. For this project, a total of 10 user stories are proposed, in Table I you can see the description of each story, which allows us to better understand the operation of the system.

B. Estimate

Here the user stories are organized through the product backlog, in this list certain criteria are taken into consideration so that the stories are properly ordered. The estimation process is carried out through planning poker, when putting this technique into practice the team must assign a number (Fibonacci series) to each of the user stories. The assigned numbers are chosen from lowest to highest according to the level of difficulty that the team members consider for the development of each user story.

In Table II, we can see the product backlog developed and ordered by priority, this is defined by the effort and difficulty of development as well as its relevance in the project. As a reference to choose the priority of each story, story number 9 has been selected as its development is considered the least difficult.

TABLE I. User History

History No.	Description
H1	As an administrator, I want a web platform that through its design is intuitive and friendly to be used easily.
H2	As an administrator, I want the web platform to allow the user to know if a seller/supplier has a history to avoid being scammed.
H3	As an administrator, I want the web platform to have a wide database to be more efficient and precise in the search for results.
H4	As an administrator, I want the platform to allow users to write reviews to better understand the type of fraud that the accused uses.
H5	As a user, I want to contribute by registering duly substantiated complaints to improve the prevention capacity of the web.
H6	As a user, I want to be able to register on the web platform in order to be prevented from various cases of fraud.
H7	As an administrator, I want the platform to have a section where information related to computer crimes is displayed to keep users informed and warned.
H8	As an administrator, I want the web system to validate the attached file that will be uploaded as evidence by the user to detect if it is a police report.
H9	As an administrator, I want the platform to have a section with external links to government pages to provide support and guidance to victims of fraud.
H10	As administrator, I want the web platform to allow knowing a ranking of fraud complaints by number of complaints and type of fraud so that users can better and easily manage the information.

TABLE II. Product Backlog

History No.	Estimate	PRIORITY	Sprint
H3	5	1	3
H8	8	2	3
H2	3	3	3
H5	3	4	2
H10	5	5	2
H4	2	6	2
H6	2	7	2
H9	1	8	1
H7	2	9	1
H1	2	10	1

The project is divided into 3 sprints, the first sprint has a total of 5 story points, because it requires less effort. When starting with the development of the sprints, the work team begins to integrate allowing better team work as each iteration progresses. The second and third sprints receive 12 and 16 story points respectively. Likewise, the points have increased due to the time and effort required by the requirements corresponding to these sprints. At this point in the project, the team is able to communicate and organize much better, thus achieving good coordination in the development of the web system and minimizing errors. Fig. 4 graphically shows the order by effort from lowest to highest of the user stories and the sprints to which they have been assigned.

C. Implementation and Development Stage

In this stage, the use of the technologies to be implemented and the procedure proposed for the web system are explained in detail.

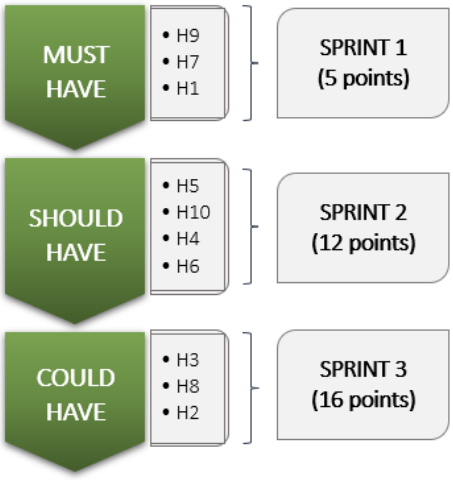

Fig. 4. Story Points and Number of Sprints.

1) Software Development: The web system was developed applying the use of the Model View Controller (MVC), which is a software architecture style based on three layers or levels. This type of model has proven its worth over several years and across various types of applications, a variety of programming languages, and development platforms. The model separates an application's data, user interface, and control logic into three distinct components:

a) Model:
Also known as the data layer, it is where the data is located and is responsible for accessing it, it is made up of one or more database managers that perform storage. It has a representation of the data used by the system, its business logic and records of the controls and views of the system.

b) View:
It is the user interface also known as the presentation layer, it shows the system to the user and interacts with it through mechanisms. It also integrates and organizes the information that is sent from the model through the controller, that is, it receives the data and shows it to the user.

c) Controller:
Also known as the business layer or business logic because it is here where all the rules that must be met are established. It acts as an intermediary between the Model and the View, managing the flow of information between them to request the database to store or retrieve data. That is, it receives requests from the user and sends responses after the process. This is also where the programs that run are located.

In Fig. 5, we can see graphically a basic scheme of how the chosen model is structured.

2) Data Set Preparation and Extraction: For the present project, a set of data was needed to be used for training and testing the different functionalities that were sought to be implemented.

In Fig. 6, we observe the process to obtain the data in image format, the social network Facebook served as the basis for extracting said information. By accessing a personal account, they entered communities called groups that the social network

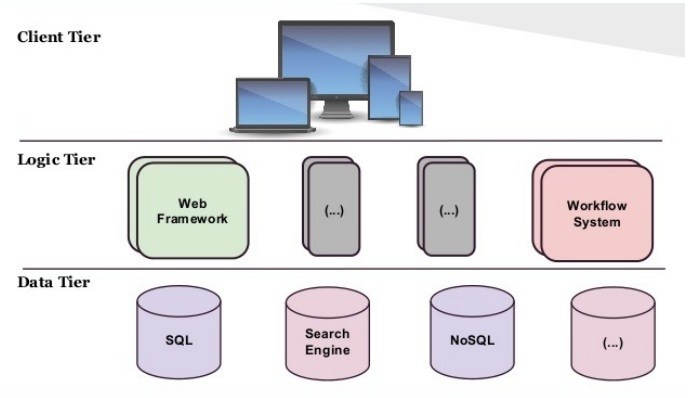

Fig. 5. Three Layer Architecture.

Fig. 6. Flowchart for Obtaining the Data Set.

offers where its users can exchange or share common interests. Then we began with the search for public publications where users of the social network made reports of having been victims of fraud.

Finally, only those publications with support were considered where images of police reports were extracted to form the data set needed by the project. This meant dividing the data into two groups, allowing training and then checking the correct functioning of the web system. In this way, it was also possible to correctly identify the types of attached files in image format that are uploaded to the system by users.

3) Optical Character Recognition (OCR):

a) Definition:
It is a software that allows us to recognize text in digital documents including images, it has become a very useful tool due to its precision, speed and efficiency to classify documents and/or extract information from them [21]. For the present research work, the Optical Character Recognition software was implemented as part of the web system for the desired function of classifying the information provided by users. In Fig. 7, we can see the process and threads carried out by the software for text extraction, once the results are obtained they can have various applications as indicated in the graph.

b) Recurrent Neural Network (RNN) and Tesseract OCR:
From simple image classification to medical scans, speech

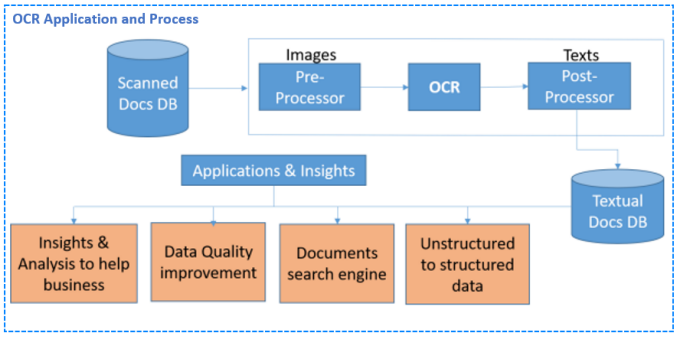

Fig. 7. Basic Flowchart of how OCR Works.

recognition and text-in-image recognition are among the many essential functions of deep learning. The RNN is a kind of neural network that is part of deep learning, it allows input data to flow in any direction along a time sequence through the layers that make up the network. The Long short-term memory (LSTM) is a type of RNN that has feedback connections, managing to process not only individual data but also complete sequences of data, avoiding loss of information over time [22].

For the proposed system, we worked with the Tesseract software, which is open source, which uses OCR and is complemented by LSTM to achieve a more efficient result in the field of text recognition through digital images. In addition, it has an extensive library that facilitates its training with a certain set of data. In Fig. 8, we can see the architecture of the Tesseract software in conjunction with the LSTM neural network.

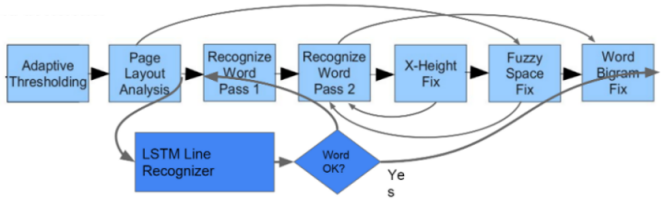

Fig. 8. Tesseract Architecture where it is Noted that there are Many Revisions of Old Decisions [23].

c) Operation:

OCR software is normally composed of four stages that are pre-processing, segmentation, character extraction and final recognition of the text, the mentioned sequence can vary depending on the results that are desired to obtain. For the present research work, since they are simple documents and in the format of a police report, the process for text recognition did not imply difficulty.

Fig. 9 shows the operation and the stages carried out by the OCR, once it manages to recognize the text in the digital document, it can also extract said information for later use, which depends on the purpose of each system that uses this technology [25].

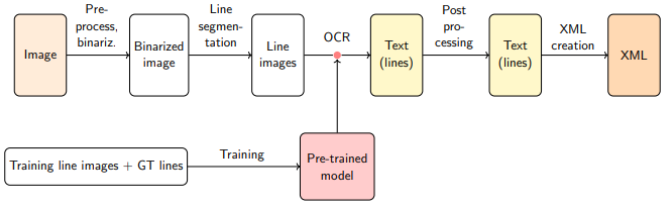

Fig. 9. Flowchart of the Process Performed by the OCR Software [24].

d) Execution and Process:

With the developed web system, the recognition software was implemented, the steps to be followed by the system as a whole were verified according to the proposal to confirm its viability and correct use. It was considered to focus a greater effort on the function of the system that requires the support of artificial intelligence because its operation is essential for the rest of the processes.

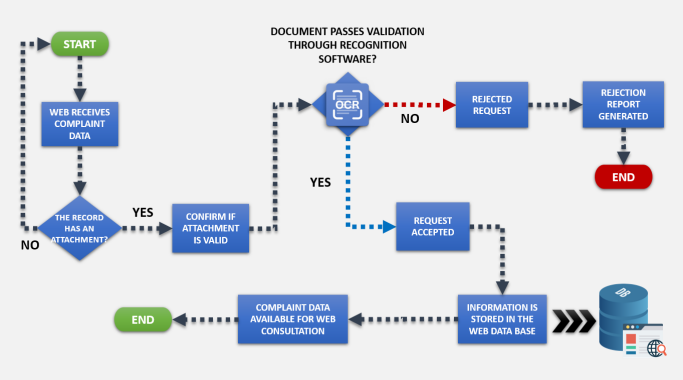

Fig. 10. Flowchart Showing Web System Working with OCR Software.

As shown in Fig. 10, the step to register a case of fraud involves attaching evidence to support it, the user through the web system must upload their respective police report in image format. Once the system receives the information of the complaint with the attached image, the recognition software will validate whether or not the image is valid by extracting the text and classifying it. If the validation result is positive, all the information entered is stored in the web system database for later use in user queries.

V. Results and Discussion

Finally, each sprint and its respective user stories are explained in detail, as well as the operation of the complete web system and we verify the importance and viability of the project through statistical data.

A. Design and Prototypes

The prototypes were developed based on user stories, these allow us to understand graphically and in detail each function implemented in the web system, as well as its importance in this project. This section shows the main prototypes organized by sprints which make up the research work carried out.

1) First Sprint: This sprint mainly covers interfaces that do not require much effort and time, with a total of three deliverables. In Fig. 11, we see the prototype that represents the home page of the web system, in it the user has the option of being able to log in or register by creating an account. You can also see a brief description of the objective of the platform and at the top is the design of the created logo.

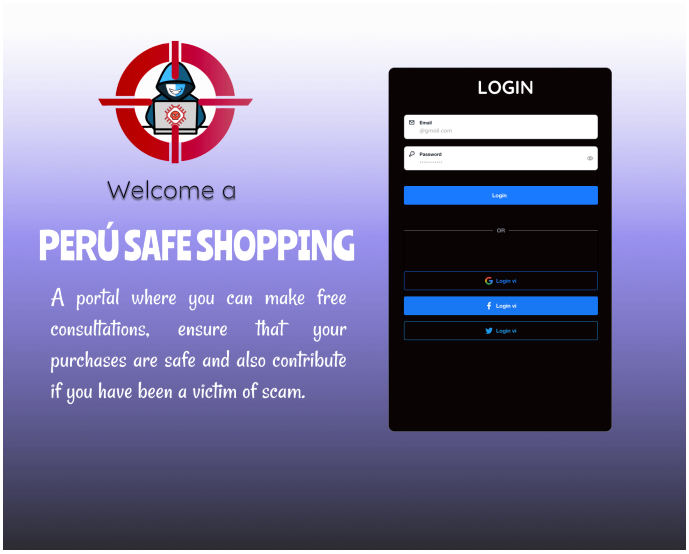

Fig. 11. Prototype based on User Story 1

In Fig. 12, we see the prototype designed for the informative section of the platform, this section includes news and advice to guide the user so that they can stay informed and prevented from the main cases of fraud and fraud at the national level.

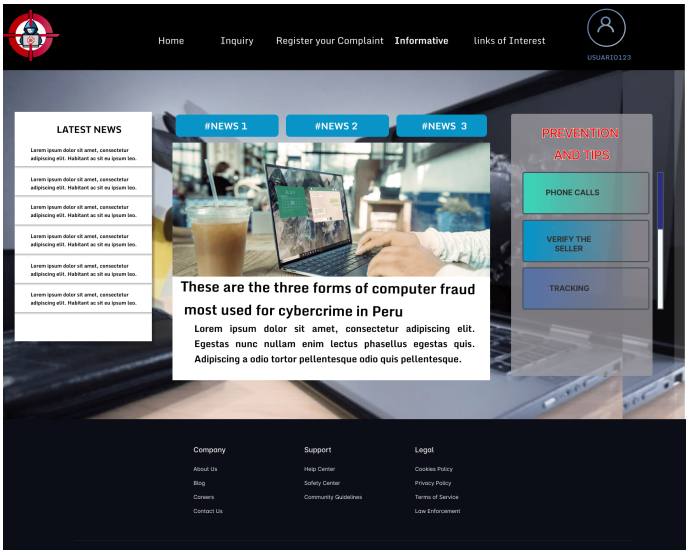

Fig. 12. Prototype based on User Story 7

2) Second Sprint: In this sprint, interfaces were mainly developed whose function is to show relevant information on fraud cases, which have been duly validated and stored in the system's database. In Fig. 13, the prototype shows the interface where the user must fill out the complaint form, an important requirement in this step is to attach the document where the police complaint is certified.

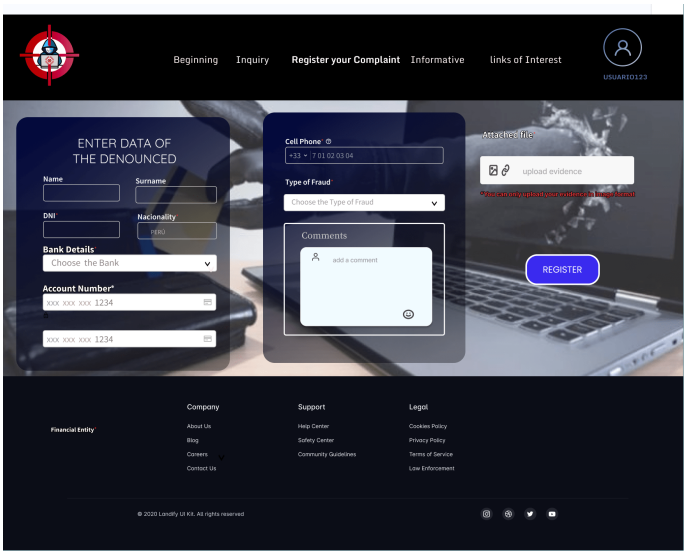

Fig. 13. Prototype based on User Story 5

The prototype in Fig. 14 shows the interface where the most reported cases of fraud can be queried, through predetermined filters such as payment method, number of reported cases and by city. In the section on the right of the interface, an easy-to-understand graph for the user is shown, where relevant information on the registered cases appears.

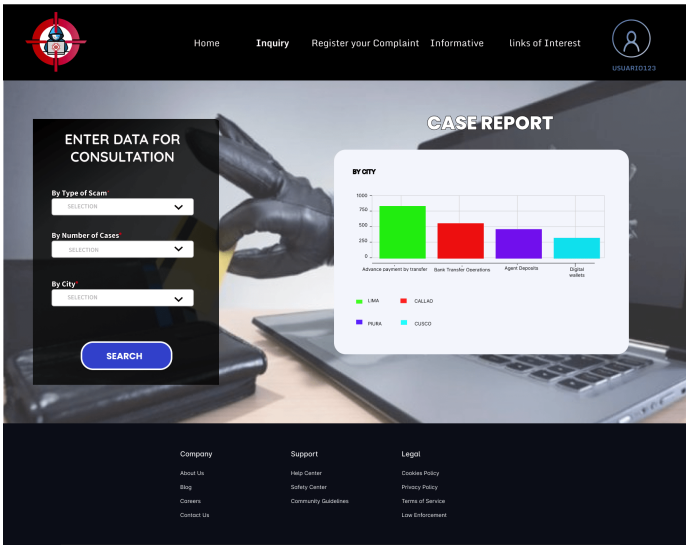

Fig. 14. Prototype based on User Story 10

In Fig. 15, we see the prototype for the user registration interface, basic data is requested for the creation of an account on the platform, also giving the alternative of being able to use a username instead of the personal name.

3) Third Sprint: This sprint is mostly related to internal processes and the logic of the web system, for that reason it is the sprint that requires the most time and effort in the

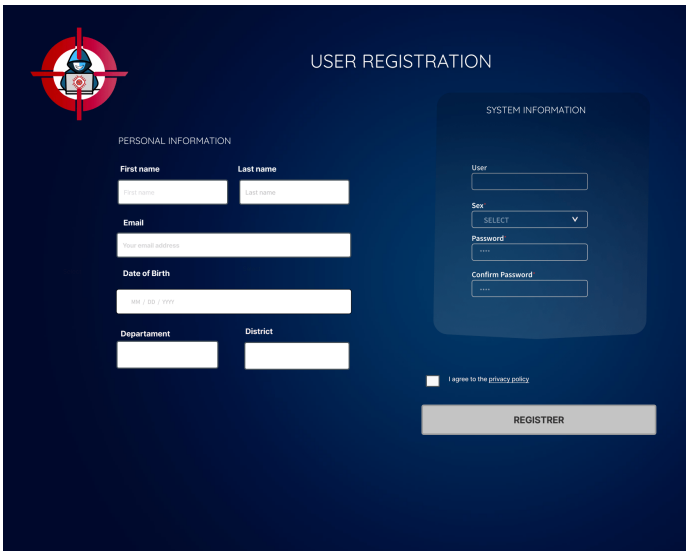

Fig. 15. Prototype based on User Story 6

development of the project. The database was initially filled manually in order to carry out the first tests of the system and detect possible errors.

Subsequently, the filling of the base continued, taking advantage of the learning process that the OCR software needed for optimal performance in the recognition of the entered documents. As mentioned in the Case Study section, the document in image format will allow the complaint record to be considered as valid or not.

Fig. 16 shows the interface where the user can check if a seller/supplier has been reported on the platform as a fraudster. Some of the requested data must be entered so that the system searches the database and displays the information of the accused in the section on the right.

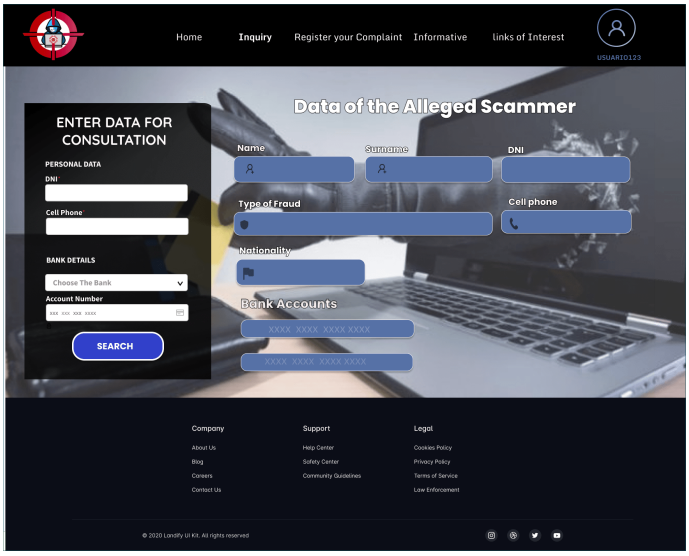

Fig. 16. Prototype based on User Story 2

B. Complaints Registered for Computer Crimes

Cybercrime in the current context is on the rise and the most affected are the consumers who have to deal with the loss of their property as a result of this type of crime. Although measures are being taken in the country and laws are being applied to regularize this situation, the efforts and policies applied have not worked properly or have not been sufficient. In Table III, we see a comparative table of the complaints made in the last five years in terms of computer crimes. Where computer fraud represents 78.2% of the reported cases, these cases are investigated by the High Technology Crimes Investigation Division (DIVINDAT).

TABLE III. COMPLAINTS OF COMPUTER CRIMES INVESTIGATED BY DIVINDAT

Crime	2015	2016	2017	2018	2019	2020	Total	%
Computer Fraud	414	610	1219	1928	2097	2615	9515	78.2
Card cloning	46	44	30	120	25	4	394	4
Fraudulent online purchases	-	-	-	287	431	261	979	8
Unauthorized electronic and/or fund transactions and transfers	368	566	1189	1521	1641	2350	8142	86

The crimes were classified according to the type of complaint: card cloning, fraudulent online purchases and unauthorized electronic and/or fund transactions and transfers, the latter being the one that forms part of this research work. The notorious increase in cases of this type of fraud was verified, registering a total of 8,142 complaints from 2015 to 2020. It was also identified that this type of fraud represents 86% of this type of computer crime, thus occupying the first place among the 3 types of cases [26].

C. Informality and Social Commerce

Informality in the country remains high, according to the Institute of Statistics and Informatics (INEI) in a study carried out in 2020. With a sample of 15,224, it was confirmed that approximately 30% of them are in the category of workers. self-employed in the informal sector [12]. With these data we can have a better picture of how broad the group of independent informal workers is, even more so if we consider that 75.3% of workers are informal at the national level, as can be seen in Fig. 17.

Informality in Peru originates mostly from tax evasion and limited resources by independent workers who fail to comply with all the requirements and procedures established to formally set up a business as required by law [12]. The limitation of resources also affects the possibility of not investing in infrastructure or having your own online store. For the aforementioned reasons, informality has also moved towards electronic commerce where sellers/suppliers establish their businesses online.

They mainly carry out their commercial activities through social networks since these do not imply any type of cost, giving greater facilities to position themselves in the digital

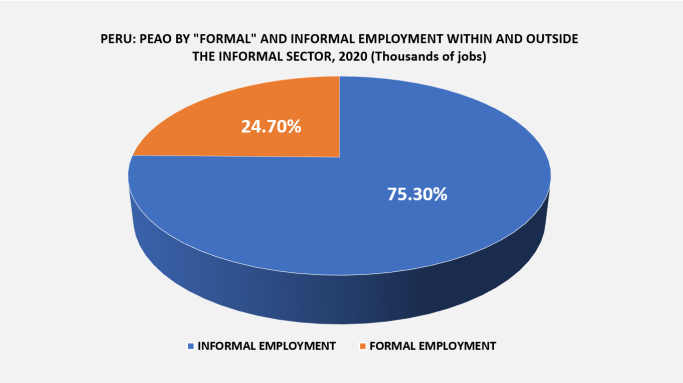

Fig. 17. The Participation of the Formal and Informal Economically Active Employed Population (PEAO) for the Year 2020 is Shown.

market. The independent worker or business owner offers their products/services describing their characteristics through Instagram or Facebook. The purchase/sale is made directly through these platforms or transferred to WhatsApp where the date and place of delivery are agreed.

D. Fraud and Social Networks

There are a large number of purchase and sale operations carried out through social networks such as Facebook, WhatsApp and Instagram where there are informal businesses with good intentions to comply with what is offered. But there are still risks for buyers since it is not the same to buy in a known establishment than to buy through a social network from a stranger. Fraudsters and/or computer criminals take advantage of the anonymity of social networks to commit their crimes, their way of operating involves creating false personal or business accounts as a facade. Then, through deception, they promise supposed products or services that they offer if they are previously paid in advance to cover the expenses involved in the delivery or fulfillment of what is offered.

TABLE IV. REQUESTS FOR ACTIVE JUDICIAL ASSISTANCE

Requirements	Number
Facebook account information	45
Email account information	18
Request information from Facebook, Messenger, WhatsApp	3
Request information from Google	2
Request information from a website	2
Arrested user identity, intervention information in the Russian social network account	1
Request information from various companies that provide social networks	1
Request removal of communications and traffic information and content of a Facebook account	1

According to the report on Cybercrime in Peru [26] prepared by the Public Ministry, 45% of active judicial assistance has been aimed at requesting information from Facebook accounts and 9% has involved social networks. In Table IV we can observe some of the requested requirements where requesting information from social networks for the cases in question stands out. These requests were made with the intention of helping the competent entities to clarify investigations and criminal proceedings that implicate social networks as a means of criminal acts related to computer fraud.

E. Risk Prevention and Minimization

Through social networks, mainly through Facebook, it was found that there is an interest on the part of citizens to make public their complaints, reports or report cases where they have been victims of fraud. However, the information through this platform is not centralized or organized, which makes access to said information difficult. In addition, through social networks there are no terms and conditions that guarantee 100% secure purchase and sale transactions. For these reasons, this project seeks to organize the data and minimize the risks due to computer fraud as a result of social commerce.

Both parties, both consumers/buyers and suppliers/sellers, can benefit from this initiative. On the one hand, the consumer/buyer who makes sure that the other party involved in the transaction has not been reported for any type of fraud. And on the other hand, the supplier/seller that manages to generate greater confidence in its clients and gain a better reputational image for its business by not being reported in the base of the proposed system.

F. Web System

The web system was designed and developed with the objective of executing a set of tasks that allow users to take preventive measures before carrying out any operation and to carry out their transactions online safely. Next, the operations of the web system are explained, grouped by their functionality, as well as the results obtained through a survey carried out on experts to validate the quality of the system.

a) Registry Functionality:
Made up of the User Registration and Complaint Registration sections, these have the function of allowing the entry and registration of data within the system. The first section mentioned authorizes access to the platform and the second section allows you to report cases of fraud by attaching your respective police report.

b) Query Functionality:
Made up of the Home, Search, Informative and Links of interest sections, these have the function of allowing the user to see information related to the topic of prevention through the different interfaces of the platform. In the particular case of the search section, it grants access to consult information on the cases reported in the system.

c) Tesseract OCR Performance:
The free software Tesseract with its LSTM-based OCR engine complemented the development of the system by facilitating the recognition of attachments provided by users; managing to identify and classify the documents as valid or not and confirming that they were a police report. The accuracy of the LSTM Recurrent Neural Network for text recognition in Urdu script was evidenced in the [27] investigation. This type of writing is based on an Arabic cursive style and due to its nature makes the text recognition process even more difficult.

However, in Fig. 18, the results obtained from this investigation are verified where the LSTM model achieves an accuracy of 98.38% despite the difficulty in recognizing the aforementioned writing. In the present research work, it was decided to use the Tesseract software precisely because of the good performance and precision that it has shown in

TABLE V. RESULT OF VALIDATION BY EXPERTS

Criterion	Questions	Media	S.D
Usability	The web system runs responsively through any type of browser.	2.87	0.35
Usability	The web system has a friendly and easy-to-use interface for users.	2.75	0.46
Usability	The web system is properly organized for users to understand.	2.87	0.35
Usability	The design of the web system works according to a programming structure.	3	0
Feasibility	The web system development budget covers cloud storage.	2.62	0.51
Feasibility	For the development of the web system there was a certain cost or budget.	2.50	0.75
Feasibility	Business strategies have been designed so that the web system generates profitability.	2.75	0.46
Scalability	The web system is stable even with the increase in user load.	2.62	0.51
Scalability	The web system has a fast response time in relation to its database.	2.87	0.35
Scalability	System processes are well established and function properly.	2.87	0.35
Innovation	The web system performs quick queries regarding information on fraud complaints.	2.87	0.35
Innovation	The web system has been created to be able to implement more information modules on computer crimes.	2.62	0.51
Innovation	The web system constantly obtains knowledge and/or relevant information.	2.75	0.46
Technology	A transactional database was used for queries and storage of information.	2.62	0.51
Technology	A specialized programming language for web systems with artificial intelligence support was used.	2.87	0.35
Technology	A framework was used for interface design and programming of the web system that is adaptable.	2.87	0.35
Technology	The web system can be adapted to any type of device.	2.75	0.46

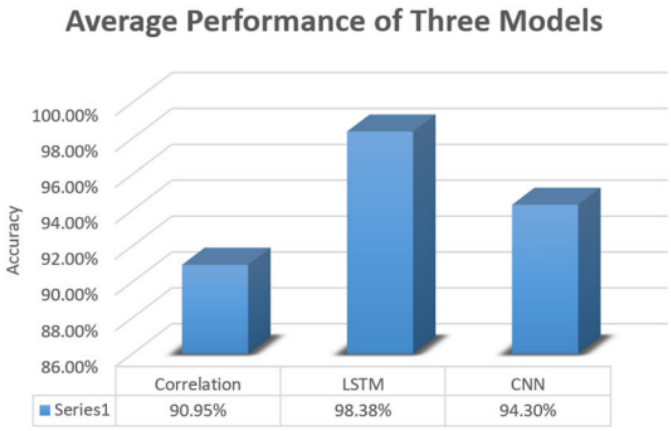

Fig. 18. Comparison between 3 Models of Artificial Neural Networks to Recognize Text in Urdu Script [27].

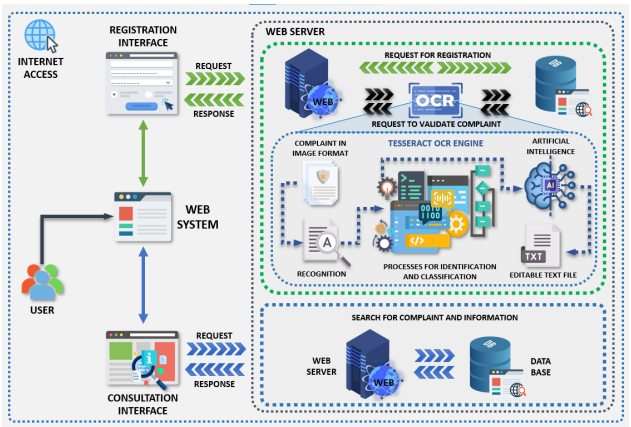

Fig. 19. Web Application Architecture.

more complex cases to identify texts in documents with image format. In addition, the software has several advantages such as having an extensive library with features that facilitate rapid learning. It also has models for several languages, covering 116 languages including Spanish, it is even recognized among the best OCR today.

d) Validation of the Proposed Model with Experts: This section specifies the results obtained through a survey of 8 experts to validate the quality level of the web system. The criteria used were: Usability, Feasibility, Scalability, Innovation and Technology. The questions raised were made using the Likert scale with the answer option: 1 (very low), 2 (intermediate), 3 (very high). The questions that were applied in the validation measure the level of acceptance of the web system by experts in the development of these applications. Table V shows the result of the validation, which is divided by different criteria as well as the proposed questions and the level of quality obtained based on the calculations for the mean and standard deviation (S.D.) of each question. The calculation of the mean allowed to establish the range of the quality level, it was obtained as a result of the total mean of all the criteria of 2.76. In that sense we can say that it has been approved by the experts.

As mentioned at the beginning of this project, the use of fraud prevention systems reduces the impact of these crimes. In the research work of [28], the favorable results were confirmed with figures when implementing a web system with the same purpose. In just 1 year, the number of computer fraud cases was reduced from 30% to 28% through its proposed model, which is also supported by artificial intelligence.

On the other hand, the investigations of [9] and [10] focus on reducing cases of fraud when purchases and payments are made through electronic commerce platforms that are part of formally constituted companies. In addition [11], [12] and [13] seek to improve the existing systems of companies that have the resources to pay for more sophisticated projects. However, the works in question do not focus or have an alternative solution for cases where purchase or sale transactions are made from person to person and that is the most used type of commerce in Peru. The proposal does not seek to favor formality, on the contrary, it seeks to offer an option for safer transactions, giving these informal businesses the opportunity and time to have enough capital to later be able to go formal.

In Fig. 19, we can see graphically the architecture of the proposed web system and the aforementioned functionalities. You can also observe in detail the support given by the OCR

tool to validate the entry of a complaint into the system, which allows the automation of this process. In this way, manual reviews of the complaints were avoided, saving time, reducing possible costs, searching for quick information and improving the management of the web system.

VI. Conclusion

Finally, the research work demonstrated how the proposed system contributes to the prevention of computer fraud caused by the digital informality present in social networks. The results of the project show how the system can reduce and minimize the risks of online purchase and sale transactions through advance payments which are made by bank transfers or the use of digital wallets. In this way, safer and more reliable operations are guaranteed in favor of the consumer/buyer who is the main affected in this new and growing digital commerce. The development of the web system aims to provide an easy-to-use tool, within the reach of the population and that allows them to be alert to situations of possible fraud. Reducing the impact caused by this type of computer crime until the competent entities can implement better measures to reinforce the computer security of citizens in the country.

VII. Future Work

As future work, it is suggested to continue investigating and incorporating different disciplines such as computer security, computer risks, cryptography. Also apply another methodology that is related to risk management. At the same time, it is recommended to delve into OCR technology to take advantage of the features it offers much more and to improve the productivity of the system by automating it. Likewise, it is possible to combine specialists in the areas of computing and electronics so that together they can contribute even more to the research project.

References

[1] World Economic Forum, *The Global Risks Report 2022 17th Edition*, 2022. [Online]. Available: https://www.weforum.org/reports/global-risks-report-2022

[2] Apoyo Consultoría, "Agenda Digital para el Perú 2021-2026," Sociedad de Comercio Exterior del Perú, Tech. Rep., 2021. [Online]. Available: https://www.comexperu.org.pe/upload/articles/publicaciones/agenda-digital-2021-2026.pdf

[3] R. V. Vereau, "Los delitos informáticos y su relación con la criminalidad económica," *Ius et Praxis*, pp. 95–110, 12 2021. [Online]. Available: https://doi.org/10.26439/iusetpraxis2021.n053.4995

[4] Cámara Peruana de Comercio Electrónico (CAPECE), "Impacto del COVID-19 en el comercio electrónico en Perú y perspectivas al 2021," CAPECE, Lima, Tech. Rep., 2021. [Online]. Available: https://www.capece.org.pe/wp-content/uploads/2021/03/Observatorio-Ecommerce-Peru-2020-2021.pdf

[5] Instituto Nacional de Estadistica e Informatica (INEI), "Producción y Empleo Informal en el Perú 2007 - 2020," INEI, Lima, Tech. Rep., dec 2021. [Online]. Available: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones-digitales/Est/Lib1828/libro.pdf

[6] C. Leyva Serrano, "Estudio de los delitos informáticos y la problemática de su tipificación en el marco de los convenios internacionales," *Lucerna Iuris et Investigatio*, no. 1, pp. 29–47, apr 2021. [Online]. Available: https//:doi.org/10.15381/lucerna.v0i1.18373

[7] F. Andrade-Chaico and L. Andrade-Arenas, "Projections on insecurity, unemployment and poverty and their consequences in lima's district san juan de lurigancho in the next 10 years," in *SHIRCON 2019 - 2019 IEEE Sciences and Humanities International Research Conference*, 2019.

[8] L. Mayer Lux and G. Oliver Calderón, "El delito de fraude informático: concepto y delimitación," *Revista Chilena de Derecho y Tecnología*, vol. 9, no. 1, p. 151, jun 2020. [Online]. Available: https://doi.org/10.5354/0719-2584.2020.57149

[9] K. Yoshida, K. Tsuda, S. Kurahashi, and H. Azuma, "Online shopping frauds detecting system and its evaluation," vol. 2. IEEE Computer Society, 9 2017, pp. 649–653. [Online]. Available: https://doi.org/10.1109/COMPSAC.2017.182

[10] K. K. M. Priyadharsini and M. S. F. I. M. Mary, "Online transaction fraud detection system," *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021. [Online]. Available: https://doi.org/10.1109/ICACITE51222.2021.9404750

[11] S. Carta, G. Fenu, D. R. Recupero, and R. Saia, "Fraud detection for e-commerce transactions by employing a prudential multiple consensus model," *Journal of Information Security and Applications*, vol. 46, pp. 13–22, 6 2019. [Online]. Available: https://doi.org/10.1016/j.jisa.2019.02.007

[12] Y. Cai and D. Zhu, "Fraud detections for online businesses: a perspective from blockchain technology," *Financial Innovation*, vol. 2, 12 2016. [Online]. Available: https://doi.org/10.1186/s40854-016-0039-4

[13] R. Jhangiani, D. Bein, and A. Verma, "Machine learning pipeline for fraud detection and prevention in e-commerce transactions," *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0135–0140, 10 2019. [Online]. Available: https//:doi.org/10.1109/UEMCON47517.2019.8992993

[14] A. Srivastava, S. Bhardwaj, and S. Saraswat, "Scrum model for agile methodology," *2017 International Conference on Computing, Communication and Automation (ICCCA)*, pp. 864–869, 5 2017. [Online]. Available: https://doi.org/10.1109/CCAA.2017.8229928

[15] F. Hayat, A. U. Rehman, K. S. Arif, K. Wahab, and M. Abbas, "The influence of agile methodology (scrum) on software project management." IEEE, 7 2019, pp. 145–149. [Online]. Available: https://doi.org/10.1109/SNPD.2019.8935813

[16] P. Ounsrimuang and S. Nootyaskool, "Introducing scrum process optimization." IEEE, 7 2017, pp. 175–181. [Online]. Available: https://doi.org/10.1109/ICMLC.2017.8107761

[17] A. Ramos-Romero, B. Garcia-Yataco, and L. Andrade-Arenas, "Mobile application design with iot for environmental pollution awareness," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, pp. 566–572, 2021, doi:10.14569/IJACSA.2021.0120165.

[18] R. A. Kacprzyk, J. Pedrycz, W. Jamshidi, M. Babanli, M. B. &. Sadikoglu, and F. M. Aliev, *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019*. Springer International Publishing, 2020, vol. 1095. [Online]. Available: http://link.springer.com/10.1007/978-3-030-35249-3

[19] G. Delanerolle, X. Yang, S. Shetty, V. Raymont, A. Shetty, P. Phiri, D. K. Hapangama, N. Tempest, K. Majumder, and J. Q. Shi, "Artificial intelligence: A rapid case for advancement in the personalization of gynaecology/obstetric and mental health care," *Women's Health*, vol. 17, p. 174550652110181, 1 2021. [Online]. Available: https://doi.org/10.1177/17455065211018111

[20] H.-H. Tseng, Y. Luo, R. K. T. Haken, and I. E. Naqa, "The role of machine learning in knowledge-based response-adapted radiotherapy," *Frontiers in Oncology*, vol. 8, 7 2018. [Online]. Available: https://doi.org/10.3389/fonc.2018.00266

[21] H. Singh and A. Sachan, "A proposed approach for character recognition using document analysis with ocr." IEEE, 6 2018, pp. 190–195. [Online]. Available: https://doi.org/10.1109/ICCONS.2018.8663011

[22] S. S. Bukhari, S. Francis, C. N. N. Kamath, and A. Dengel, "An investigative analysis of different lstm libraries for supervised and unsupervised architectures of ocr training," vol. 2018-August. IEEE, 8 2018, pp. 447–452. [Online]. Available: https://doi.org/10.1109/ICFHR-2018.2018.00084

[23] S. M. Shithil, A. R. M. Kamil, S. Tasnim, and A. A. M. Faudzi, "Container iso code recognition system using multiple view based on google lstm tesseract," in *Computational Intelligence in Machine Learning*. Springer, 2022, pp. 433–440. [Online]. Available: https://doi.org/10.1007/978-981-16-8484-5_41

[24] S. Drobac and K. Lindén, "Optical character recognition with neural networks and post-correction with finite state methods," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 23, pp. 279–295, 12 2020. [Online]. Available: https://link.springer.com/10.1007/s10032-020-00359-9

[25] Hubert, P. Phoenix, R. Sudaryono, and D. Suhartono, "Classifying promotion images using optical character recognition and naïve bayes classifier," *Procedia Computer Science*, vol. 179, pp. 498–506, 2021. [Online]. Available: https://doig.org/10.1016/j.procs.2021.01.033

[26] Ministerio Público Fiscalia de la Nación, "Ciberdelincuencia en el perú: Pautas para una investigación fiscal especializada," pp. 1–70, 2021. [Online]. Available: www.mpfn.gob.pe

[27] A. Naseer and K. Zafar, "Meta features-based scale invariant ocr decision making using lstm-rnn," *Computational and Mathematical Organization Theory*, vol. 25, pp. 165–183, 6 2019. [Online]. Available: http://link.springer.com/10.1007/s10588-018-9265-9

[28] J. C. Moreno, C. M. S. M. Sánchez, J. Salavarrieta, and L. M. Vargas, "Soluciones tecnológicas para la prevención de fraude y diseño de un modelo de prevención del riesgo transaccional para el botón de pago," *Entre Ciencia e Ingeniería*, pp. 36–42, 12 2019. [Online]. Available: https://doi.org/10.31908/19098367.1154