

**Integrated Project Data Science**

**Development Of an Autocorrection System of Bank Checks**

***Elaborated by:***

**Houssem Ouerdiane**

**Haithem Gharbi**

**Youssef Mansour**

**Aicha Chaaba**

**Molka Mrad**

**Dhia Ben Amor**

Supervised By:

|  |  |
| --- | --- |
| Mme. Sonia Mesbah | Mme. Dorra Trabelsi |

Academic Year

**2021 – 2022**

Table Of Contents

Chapter I. Understanding the Concept & Identifying the Problematic

I.1 Introduction

I.2 The Problematic

Chapter II. Identifying and Explaining the functional needs

Chapter III. Identifying and Explaining the non functional needs

III.1 Artificial Intelligence

III.2 Computer Vision

III.3 Natural Language Processing

Chapter IV. Identifying and Explaining the Constraints concerning the development of the Product

Chapter V. Conclusion

List of abbreviations

ACH: Automatic Clearing House

EFTPOS: Electronic Funds Transfer At Point Of Sale

ATM: Automated Teller Machines

CTS: Cheque Truncation Systems

AI: Artificial Intelligence

NLP: Natural Language Processing

OCR: Optical Character Recognition

CNN: Convolutional Neural Network

MNIST: Modified National Institute of Standards and Technology

EMNIST: Extended Modified National Institute of Standards and Technology

SVM: Support Vector Machine

SIFT: Scale Invariant Feature Transform

**Chapter I. Understanding the Concept & Identifying the Problematic**

**I.1 Introduction**

Since the establishment of banks, a financial revolution has changed our lives for the better.

In fact, Banks play a very important role in the economy of any country and as a citizen it offers you many advantages.

There is no doubt that they offer a number of useful services, it offers customers different payment methods such as Automated Clearing House (ACH), Wire transfers or telegraphic transfer, EFTPOS, automated teller machines (ATMs) and obviously checks.

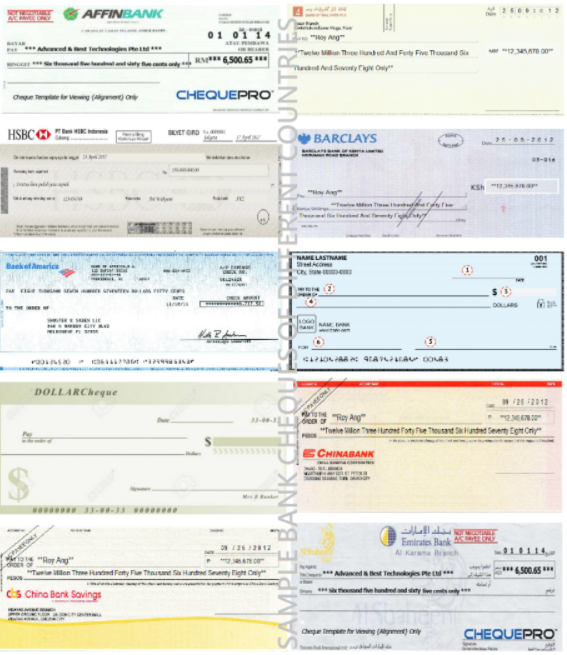
**I.2 The Problematic**

In every country, banking institutions have provision for their customers to use checks for any financial transaction.

Check systems in banks for transactions is one of the most powerful tools, however, due to its slow processing time, considerable complexities in check clearing process and gradually with an immense growth in the database, the electronic-transactions have created a big shadow on the checks system to cater the needs of the banking sector.

In a manual verification, user written information including date, signature, legal and courtesy amounts present on each cheque has to be visually verified. As many countries use cheque truncation systems (CTS) nowadays, much time, effort and money can be saved if this entire process of recognition, verification and data entry is done automatically using images of cheques.

Those are examples of different checks from different countries:



As you can see, there are different types of checks, each check has different aspects such as format of writing, dimension, colors, police, language...

For those reasons, it becomes difficult for an automated system to identify the contents of these cheques. Moreover, human mistakes are also possible in these cases, a person may write something down in a check that is not valid or contains false information

let’s take the example of someone writing down an amount that contains:

- Non-matching amount in letters and in numbers

- Syntax mistakes

- Invalid date

- Forgetting to fill a part of the check

- Cross-outs

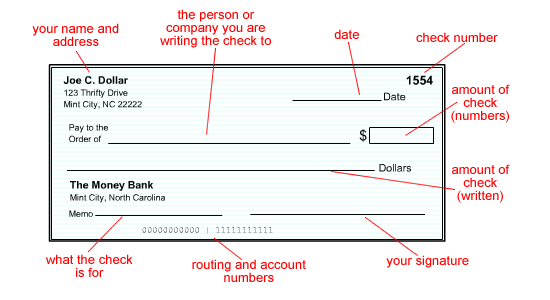
- Unreadable handwriting

**Chapter II. Identifying and explaining the functional needs**

We have proposed a mechanism to make the entire process of bank check verification not only fast but also reliable by using deep learning methods and image processing techniques.

Those are our main objectives:

* Achieve the autocorrection of the amount in a check
* Manipulate a huge amount of data
* Assure the security in the process
* detection of the different parts of the cheques
* Comparison between the amount in numbers and the amount in letters
* verify the validity of the date (the check expires after 3 years and a week)
* Check if the signature exists and is valid
* Check which language is used in the check
* The ability to identify different handwritings



**Chapter III. Identifying and explaining the non-functional needs**

**III.1 Artificial Intelligence**

AI is the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages.

**III.2 Computer Vision**

Computer vision is a field of artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs — and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand.

*Preprocessing:*

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subfield of digital signal processing, digital image processing has many advantages over analogue image processing. It allows a much wider range of algorithms to be applied to the input data — the aim of digital image processing is to improve the image data (features) by suppressing unwanted distortions and/or enhancement of some important image features so that our AI-Computer Vision models can benefit from this improved data to work on.

*Segmentation:*

Capturing the components of a scanned bank check image using image segmentation: Date, Signature of account holder, Courtesy amount detail (numbers), legal amount detail (letters).

**III.3 NLP (Natural Language Processing)**

NLP is the field of artificial intelligence that relates lingual to Computer Science.

- Identify Misspelled Word

- Find distance away

- Filtering

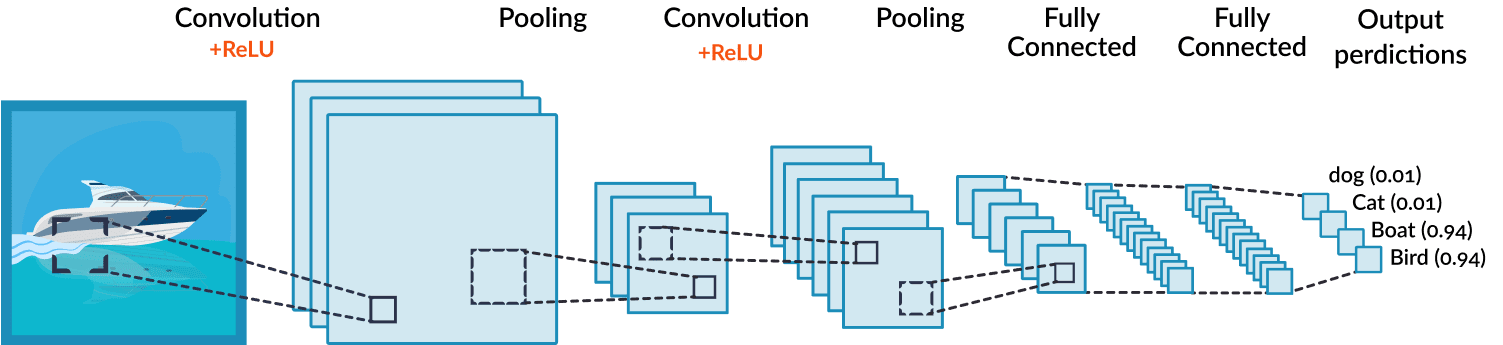
- Calculate Probabilities of Words

*OCR (Optical Character Recognition):*

For Legal amount and Courtesy amount identification OCR can be implemented by using Convolutional Neural Network (CNN), which is a popular deep neural network architecture.

Identifying the digits from courtesy amount image segment and characters from legal amount image segment using CNN, EMNIST dataset for Latin letters and MNIST dataset for numbers.

->The EMNIST Dataset is an extension to the original MNIST dataset to also include letters.



*SVM (Support Vector Machine) & SIFT (Scale Invariant Feature Transform):*

For verification of Signature, we have used Scale Invariant Feature Transform (SIFT) for extraction of features and Support Vector Machine (SVM) as classifier with a dataset that contains signatures named BANGLA.

The SIFT allows you to compare stored features of an original signature to copies and identify a match with a high level of accuracy even in the cases that the sample is scaled or rotated.

**Chapter IV. Identifying and Explaining the Constraints concerning the development of the Product**

* **Confidentiality of bank clients**

The bank's duty of confidentiality covers all customers' information about themselves and their accounts obtained by the bank, irrespective of the information source and for as long as the banker-customer relationship exists.

There are many logical reasons for obliging the banks to keep customers' confidential data private before, during, and after their relationship which makes the collect of the data difficult.

* **The diversity of checks layouts**

Automatic bank check processing is a very difficult task due to the diversity of the check’s layout.

In fact, the emplacement of check fields differs from one country to another which makes the segmentation quite hard.

* **Budget limits**

Since we are still students, we don’t have a proper funding for our project. We must make do with our limited resources.

* **The Security of the system**

Since our system works on bank checks we must assure that the whole process is secure and protected from any kind of threats or attacks.

* **High-Availability of the server**

In order to assure the rapidity of the process, we must make sure that the web server in which we will deploy our project is always accessible and available for the bank agents.

* **Reduce the amount of paper used in checks**

In the USA, paper checks are on the decline. Back in 2000, paper checks were used in 42.6 billion transactions.

In 2018, that number significantly dropped to 14.5 billion.

**Chapter V. Conclusion**

To sum up the whole process, here is this scheme:

