EKYC Mobile Application using Optical Character Recognition

Yash Kumar
Department of Computer Engineering
Terna Engineering College
Nerul, Navi Mumbai, India

Komal Sakpal
Department of Computer Engineering
Terna Engineering College
Nerul, Navi Mumbai, India

Gaurav Sharma
Department of Computer Engineering
Terna Engineering College
Nerul, Navi Mumbai, India

Prof. Audumbar Umbare
Department of Computer Engineering
Terna Engineering College
Nerul, Navi Mumbai, India

Abstract— **The purpose of this paper is to create an android mobile application with the help of Machine Learning using Optical Character Recognition for the ease of customers to update and validate their credentials for banking purposes like creating a bank account, applying for a loan, and various other business purposes, for the purpose of convenience and saving time. This will not only speed up the process of KYC phenomenally, it will also make it error free. In our mobile application, the customers can update their KYC by capturing photos of their AADHAR and PAN credentials. The app will use OCR (Optical Character Recognition) to minimize the typing errors and thus fills up the form without errors, saving valuable time.**

Keywords— *Android, KYC, Optical Character Recognition, Image Processing, credentials.*

I. INTRODUCTION

KYC means "Know Your Customer". It is a process by which banks obtain information about the customer's identity and address. This helps ensure that bank services are not misused. The KYC procedure is to be completed by banks while opening accounts, and also, to periodically update the same. To open a bank account, one needs to submit Aadhaar/enrolment number and PAN as 'proof of identity and proof of address' together with a recent photograph.

To create an EKYC application to prevent the wastage of forms during KYC process. When a customer goes to a bank to open an account or to update his/her KYC (Know-Your-Customer), he/she has to fill different forms and needs to take Xerox copies of different documents. Initial process when a customer visits a bank is to fill forms and submit copies of the documents. Many a times there is a risk in manual errors, leading to the wastage of papers, as well as time. In our mobile application, the customers can update their KYC by capturing photos of their AADHAR and PAN credentials. The app will use OCR (Optical Character Recognition) to minimize the typing errors and thus fills up the form without errors, saving valuable time.

II. PURPOSE

Normally, customers fill forms manually on sheets of paper. This can lead to a lot of discrepancy because of human errors, illegible handwriting, and malfunctioning writing material. This could lead to a lot of wastage of paper (considering this is a process adopted worldwide). Next, there could be errors made by the authorities responsible for data entry, while referring the handwritten form. Further, there could be inconsistency in the customer data because of the way customers enter their data, causing major validation problems to the customer whose original purpose was to validate their identity in the first place. The answer is to use credentials recorded by recognized authorities like Income Tax Department and the Unique Identification Authority of India (UIDAI), such as PAN card and Aadhaar Card, to help solve this problem. These official credentials are used to enter customer data for KYC by capturing user data from the physical card.

III. OBJECTIVES

Following are the objectives of the mobile application:
• Scanning Credentials like PAN Aadhaar, Driving License, and Passports with the help of Optical Character Recognition.
• Implementing support for different sizes and formats of each of these credentials to prevent application failure.
• Identifying and collecting User Information using these credentials.
• Creating a database for of the information collected for each customer.
• Filling up customer details in forms demanding their information.
  A few goals of this project are defined:
• Driving down cost of operations by detecting inefficiency in customer identification.
• Reducing risks of identity theft.
• By increasing customer satisfaction in a number of different functional operations across the institution.
• Improving overall quality of life of customers as well as verification officials

IV. LITERATURE REVIEW

KYC (Know Your Customer) is becoming a critical gatekeeper process for financial institutions, the world over, to safeguard against financial frauds, terrorist funding and money laundering. It involves collecting basic identity & address information about the customer. [1] One has to submit some documents to authenticate identity and address of the

client/customer. The list of documents required is mentioned below:-
  • Proof of Identity
  • Proof of Address

KYC process is and has been carried out in the following ways:-

1. KYC Offline: A customer can do KYC offline as well. However, it may take up to 7 days for the KYC to be approved by the KRA (KYC Registration Agency). You have to follow the steps mentioned below for doing KYC offline:
  • Download and fill the KYC form
  • Mention your Aadhaar/PAN details
  • Visit a KRA office and submit the application
  • Attach the proof of identity and proof of address with the application
  • You may have to submit your biometrics as well in some cases
  • You will get an application number which can be used to check the status of the KYC

2. KYC Online: Aadhaar OTP allows one to get the KYC done quite easily in minutes. You have to follow the steps mentioned below for doing KYC online:
  • Visit the website of any KRA (KYC Registration Agency)   or a fund house
  • Some of the KRAs are as follows – NDML, CAMS, Karvy, CVL and NSE
  • Enter your details as mentioned in your Aadhaar card.
  • Verify using the C where you have to enter the OTP sent to the mobile number registered with Aadhaar.
  • Submit your application
  • Once verified with UIDAI, the KRA approves your KYC
  • You can check the status of your KYC request by visiting the portal of the KRA using your PAN

3. KYC Online Aadhaar-based Biometric: Aadhaar-based Biometric KYC, one has to apply for KYC online and an executive from the KRA visits his home/office for biometric verification. You have to follow the steps mentioned below for online KYC using Aadhaar Biometric Authentication:
  • Visit the portal of any KRA or fund house
  • Perform online KYC as mentioned in the process above
  • Request for biometric authentication online
  • An executive from the fund house visits the address mentioned in the form
  • Show him your original documents and provide your biometrics
  • Your application will be submitted and KYC will be done

The key to survival in today's financial services market can be summed up as: "Better know your customer." The International Journal of Computer Applications (0975 – 8887) Volume 97– No.9, July 2014 50 identification of a customer is a very critical process in KYC with a view to protect the customer interests by preventing from fraudsters who may use the name, address and forge signature to undertake illegal business activities, encashment of stolen drafts, cheques, etc. This also helps to safeguard banks from being unwittingly used for the transfer of funds derived from criminal activity or for financing terrorism. Identification of customers also helps in controlling financial frauds, identify money laundering and suspicious activities, and for scrutiny / monitoring of large value cash transactions.

In India, in order to prevent these issues, the Reserve Bank of India (RBI) had directed all banks and financial institutions to put in place a policy framework to know their customers before opening any account. This involves verifying customers' identity and address by asking them to submit documents that are accepted as relevant proof. Mandatory details required under KYC norms are proof of identity and proof of residence. Passport, Voter's ID card, Permanent Account Number (PAN) card or driving license are accepted as proof of identity, and proof of residence can be a ration card, an electricity or telephone bill or a letter from the employer or any recognized public authority certifying the address, in addition to proof of identity being used as residence proof in case they carry address.

More recently Aadhaar card (Unique Identification Number to all Indian citizens given by Unique Identification Authority of India (UIDAI)) is being used as a valid KYC document as both proof of identity and proof of address. Recent advancements have brought about EKYC (Electronic– Know Your Customer) using Aadhaar where only biometrics are provided and identity & address is verified online. Some banks may even ask for verification by an existing account holder. Though the standard documents that are accepted as proof of identity and residence remain the same across various banks, some deviations are permitted, from bank to bank. Similarly most high value financial transactions require customers to disclose their PAN numbers. [1] [2]

The problem with the previous implementations of KYC process is that they involve a lot of human contact and interaction with the entry data. This results in errors that cannot be foreseen, and which may persist if not corrected and accounted for.

• With respect to offline method, customers fill forms manually on sheets of paper. This can lead to a lot of discrepancy because of human errors, illegible handwriting, and malfunctioning writing material. This could lead to a lot of wastage of paper (considering this is a process adopted worldwide). Next, there could be errors made by the authorities responsible for data entry, while referring the handwritten form. [3]

• With respect to online method, consisting of OTP authentication, while it is very secure, the same problems like in offline method stand true. Typing in inputs in fields may sometimes encounter data entry errors, making it difficult during verification. Further, there could be inconsistency in the customer data because of the way customers enter their data, causing major validation problems to the customer whose original purpose was to validate their identity in the first place. [4]

• Lastly, with regards to the Aadhaar-based Biometric process, while it sounds good and convenient and easy, having a personnel come over to your house to verify credentials, it is required that you stay present at home the whole day he is to come for verification. It is also very time consuming.

To be rid these problems, we have introduced our EKYC Mobile Application, which scans, and directly fills the details after recognizing the characters in credentials like Aadhaar and PAN, reducing errors, as well as effort to type. It also authenticates the credentials, leaving no doubt about security. Since all this happens instantly, it is also saves a lot of time.

V.   METHODOLOGY

We have authorized the user for which we have created a dummy database of Aadhaar cards, with an Email Address to send an OTP (One Time Password) to it. We are creating an official Email Address, from which an authentication mail will be sent to the user, to OTP to the email specified in the database. After the OTP verification has completed, the user is given utility to take photos of their PAN cards and Aadhaar cards.

We have used OpenCV to create a Scanner which uses Laplacian Sharpness algorithm, and we have used 8-point cropping to crop skewed images. The scanned images are stored in the phone memory cards and then we use Text Recognizer to read the text of the Aadhaar and Pan Card but we have take only selective texts so we use Java Regex class to learn the pattern that occur while writing a "Name", "PAN Card Number", "Aadhar number".

A few relevant classes and methods used from OpenCV are:-
1. opencv.core.Core
2. opencv.core.Mat
3. opencv.core.MatOfInt
4. opencv.core.MatOfPoint
5. opencv.core.MatOfPoint2f

A few relevant classes and methods used from Google Vision API are:-
1. google.android.gms.vision.Frame
2. google.android.gms.vision.text.TextBlock
3. google.android.gms.vision.text.TextRecognizer

A few Regex patterns for each case are as follows:
1. Date:^([0-2][0-9]|(3)[0-1])(\/)(((0)[0-9])|((1)[0-2]))(\/)\d{4}$
2. PAN no: ^([a-zA-Z]){5}([0-9]){4}([a-zA-Z]){1}?$
3. Full name: ^([a-zA-Z0-9]+|[a-zA-Z0-9]+\s{1}[a-zA-Z0-9]{1,}|[a-zA-Z0-9]+\s{1}[a-zA-Z0-9]{3,}\s{1}[a-zA-Z0-9]{1,})$
4. Aadhaar no: ^\d{4}\s\d{4}\s\d{4}$

The final part is the address, and since we are taking address as a whole we need not use regex to identify. After done with this we have to upload the image of the person who is updating his/her KYC a photo which is given for passport and a signature or the person the signature photo is shrunk in size and both the photos are stored in the form Base64 format. We have used XAMPP to create a local host server for both dummy database and for database storing the details of the customer after the KYC is done (bank database). The complete workflow is carried out in an Android mobile application, built using Android Studio. [5] [6] [7]

VI.   FEASIBILITY
Following is the analysis of feasibility for the project:-

• Technical Feasibility: Our project is technically feasible, since we have all the hardware and software required to implement our project. (Laptops with decent processing power, Android smart phones to test our application, MATLAB and Android Studio)
• Economic Feasibility: Our project is an economically feasible product since it has nearly no cost, as it's a student product.
• Operational Feasibility: It consists of how feasible the product is for the user, how it satisfies their needs, and how 'usable' it is for a base user, with minimal knowledge of using applications.
• Scheduling Feasibility: It refers to the feasibility of the time required to complete the project within allotted/available time and whether the final product will be of use once completed, rather than some other better technology being developed. [8]

VII.   CONCLUSION
For any business or financial venture, service providers have a requirement and need to identify their customers and understand the risks they are prone to, before providing services. When prospective customers lack formal identification, or when their identification is difficult to authenticate, providers cannot easily verify their identities or perform customer due diligence (CDD) on them. EKYC is a process in which approved entities query a digital (and usually national) ID system to authenticate or verify their customers' identities and, in some cases, retrieve basic information about them.

Our EKYC mobile application improves the registering process by reducing/eliminating paper-based procedures and record-keeping, which in turn reduces cost and time spent on verification, making it easier to provide credential services to low-income customers. Use of Optical Character Recognition (OCR) techniques to accurately decipher scanned documents allows reduction in errors compared to all previous methods of performing the task. Allowing customers to electronically provide their demographic and personal information including proof of identity, proof of address, date of birth and gender to the bank which can verify it in real time, by themselves also feels empowering, and makes the customer feel as though they are in control. [9]

REFERENCES

[1] Soni, Anuraj and Reena Duggal. "Reducing Risk in KYC (Know Your Customer) for large Indian banks using Big Data Analytics." (2014)

[2] What is KYC and How to do it? ~ paisabazaar.com (faq)

[3] Why Aadhar Paperless Offline e-KYC – Aadhaar Paperless Offline E-KYC

[4] Aadhaar-based e-KYC to boost user adds for telcos: Experts - Economic Times

[5] Everything you need to know about Building an OCR Scanner from Scratch – infrrd.ai (blog)

[6] OpenCV Java Documentation (4.2.0-dev)

[7] Text Recognition API Documentation

[8] Impact of Agile Methodology on Software Development Process – Gaurav Kumar, Pradeep Kumar Bhatia

[9] Overcoming the "Know Your Customer" Hurdle with E-KYC - Michael Pisa and Jim Woodsome