International Journal of Research in Engineering, Science and Management
Volume-3, Issue-7, July-2020
journals.resaim.com/ijresm | ISSN (Online): 2581-5792

DocFace+: ID Document to Selfie Matching Using Blockchain

G. S. Akshaya[1*], R. Harshitha[2], P. Preethi[3], S. Kumari[4]
[1,2,3]*Student, Department of Information Technology, Panimalar Engineering College, Chennai, India*
[4]*Assistant Professor, Department of Information Technology, Panimalar Engineering College, Chennai, India*
*Corresponding author: gsakshaya1@gmail.com

***Abstract*: Varied activities in our manner needs us to verify who we are by showing our ID documents containing face pictures like passports and driver licenses, to human operators. However, this method is slow, labor intensive and unreliable. As such, AN automatic system for matching ID document photos to live face pictures (selfies) in real time and with high accuracy is needed. Throughout this paper, we have a tendency to propose DocFace+ to satisfy this objective. We have a tendency to 1st show that gradient-based improvement strategies converge slowly (due to the under fitting of classifier weights) once several categories have solely some samples, a characteristic of existing ID-selfie datasets. To beat this defect, to update the classifier weights, that permits quicker convergence and additional generalizable representations. Next, a combine of relation networks with part shared parameters square measure trained to hunt out a unified face illustration with domain-specific parameters. Cross-validation on AN ID selfie dataset shows that whereas a publically on the market general face intermediary.**

***Keywords*: ID-selfie face matching, Face recognition, Face verification, Access management, Document photos, Selfies.**

1. Introduction

In four years of the analytics designed quite eightieth of classification models and simply 15-20% regression models. These ratios square measure typically additional or less generalized throughout the trade. The principle of a bias towards classification models is that the bulk analytical drawback involves creating an alternative. As AN example can a client attrite or not, ought to we have a tendency to target client X for digital campaigns, whether or not client options a high potential or not etc. This analysis is additional perceptive and directly links to an implementation roadmap. throughout article, we'll mention another wide used classification technique referred to as K-nearest neighbors (KNN). Our focus are becoming to be whole on however will the rule work and so the manner will the input parameter result the output.

2. System Analysis

A. Existing system

In the existing system, biometric identification plays a vital role in our daily lives. For instance, access management, physical security and international border crossing needs us to verify our security level and our identities to verify United Nations agency we tend to area unit by showing our ID documents containing face pictures, like passports and driver licenses, to human operators. However, this method is slow, labor intensive and unreliable. As such, an automatic system for matching ID document photos to measure face pictures (selfies) in real time and with high accuracy is needed. when validatory a traveler's identity by face comparison, the gate is mechanically opened for the individual to enter. For IDselfie matching, they're scrutiny a scanned or digital document picture.

B. Disadvantages of existing system

- The downside of ID-selfie matching poses varied challenges that are totally different from general face recognition. For typical free face recognition tasks, the most challenges are thanks to cause, illumination and expression variations.
- The calibre of document photos thanks to image compression and the massive time gap between the document issue date and therefore the verification date stay because the primary difficulties.

C. Proposed system

We are proposing a certificate system supported blockchain to beat the matter. Information are hold on in several nodes, and anyone World Health Organization needs to switch a specific internal data point should request that different nodes modify it at the same time. Thus, the system is extremely reliable. we tend to developed a suburbanised application and designed a certificate system supported Ethereum blockchain. This technology was hand-picked as a result of it's incorrupt, encrypted, and traceable and permits information synchronization. By group action the options of blockchain, the system improves the potency operations at every stage. The system saves on paper, cuts management prices, prevents document forgery, and provides correct and reliable info on digital certificates and compare user live face with verified document face.

D. Advantages of proposed system

It solves the matter of counterfeiting certificates. We have a

tendency for proposing associate digital certificate system supported block chain technology and to verify the traveler's identity with live camera, that permits quicker convergence and a lot of generalizable representations.

E. Software requirements

- Tomcat 7.0
- MySQL
- JDK 1.7
- J2EE
- Windows 7 and above

F. Hardware requirements

- RAM : 4GB and Above
- Processor : P IV and Above
- Hard Disk : 80GB and Above

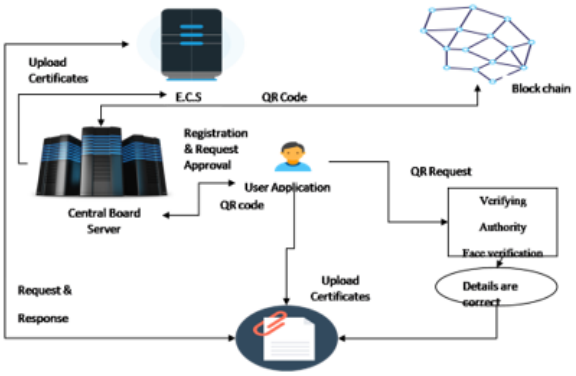
G. Technologies used

- J2EE (JSP, Servlet), JavaScript, HTML, CSS, AJAX.
- Hibernate Framework
- MVC Pattern
- Design Pattern

3. System Design

A. System Architecture

The user would register and request for the issuing of the passport. The central board server would check the small print and send the approval if everything is correct. Then the user got to upload the certificates and would request for a QR code. After the verification of certificates, authority and face recognition when the small print are correct a QR code are going to be generated. With the QR code we will access our certificates anytime.



4. System Implementation

Algorithms used:

- KNN
- SHA-256
- RSA

A. KNN algorithm

KNN are often used for each classification and regression prophetic issues. However, it's a lot of wide utilized in classification issues within the business. to judge any technique, we have a tendency to typically inspect three vital aspects:

1. Calculate time
2. Ease to interpret output
3. Predictive Power

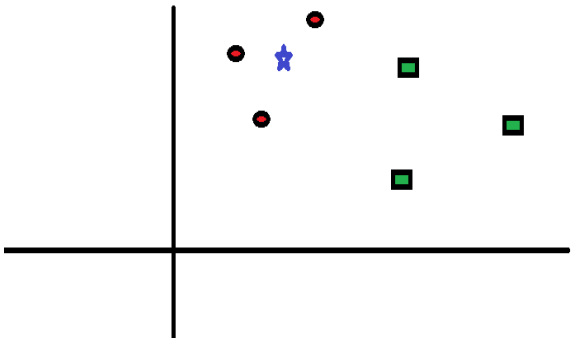
Let us take many examples to position KNN within the scale:

	Logistic Regression	CART	Random Forest	KNN
1. Ease to interpret output	2	3	1	3
2. Calculation time	3	2	1	3
3. Predictive Power	2	2	3	2

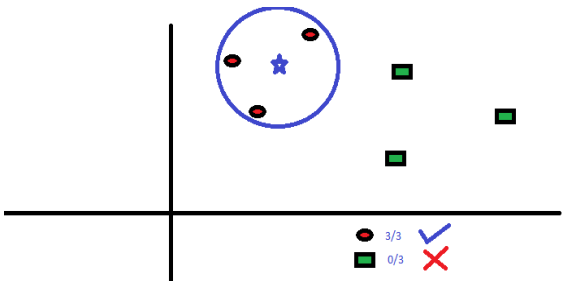
KNN rule fairs across all parameters of concerns. It's usually used for its straightforward of interpretation and low calculation time.

The KNN algorithm work

Let's take a straightforward case to grasp this algorithmic rule. Following may be an unfold of red circles (RC) and inexperienced squares (GS):



You intend to seek out the category of the suffrutex (BS). BS will either be RC or GS and zip else. The "K" is KNN algorithmic rule is that the nearest neighbors we have a tendency to want to require vote from. Let's say K = three. Hence, we are going to currently build a circle with BS as center even as massive on enclose solely 3 knowledge points on the plane. sit down with following diagram for a lot of details:



The 3 nearest points to BS is all RC. Hence, with smart confidence level we are able to say that the BS ought to belong to the category RC. Here, the selection became terribly obvious

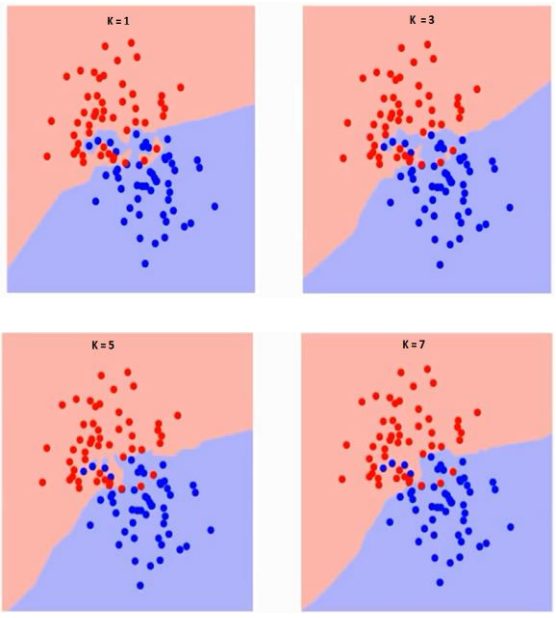
as all 3 votes from the nearest neighbour visited RC. the selection of the parameter K is extremely crucial during this algorithmic rule.

The KNN Algorithm

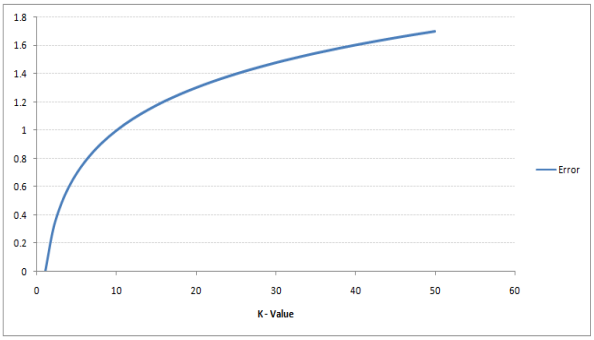
- Load the information
- Initialize K to your chosen variety of neighbors
- For every example within the knowledge
- Calculate the space between the question example and therefore the current example from the information.
- Add the space and therefore the index of the instance to AN ordered assortment
- Sort the ordered assortment of distances and indices from smallest to largest (in ascending order) by the distances
- Pick the primary K entries from the sorted assortment.
- Get the labels of the chosen K entries
- If regression, come the mean of the K labels
- If classification, come the mode of the K labels

How do we choose the factor K?

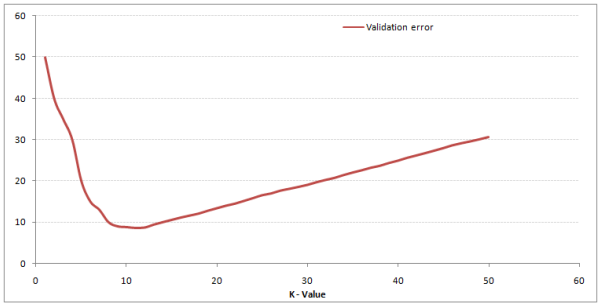
First allow us to try and perceive what specifically will K influence within the algorithmic program. If we have a tendency to see the last example, only if all the half dozen coaching observation stay constant, with a given K price we will create boundaries of every category. These boundaries can segregate RC from GS. Identical manner, let's try and see the impact valuable "K" on the category boundaries. Following are totally different the various} boundaries separating the 2 categories with different values of K.



If you watch rigorously, you'll be able to see that the boundary becomes drum sander with increasing price of K. With K increasing to time it finally becomes all blue or all red reckoning on the overall majority. The coaching error rate and therefore the validation error rate are 2 parameters we want to access on completely different K-value. Following is that the curve for the coaching error rate with varied price of K:



As you'll be able to see, the error rate at K=1 is usually zero for the coaching sample. this can be as a result of the nearest purpose to any coaching information is itself. Therefore, the prediction is usually correct with K=1. If validation error curve would be similar, our alternative of K would are one. Following is that the validation error curve with varied price of K:



This makes the story additional clear. At K=1, we have a tendency to were over fitting the boundaries. Hence, error rate ab initio decreases and reaches a lowest. Once the minima purpose, it then will increase with increasing K. to urge the optimum price of K, you'll be able to segregate the coaching and validation from the initial dataset. Currently plot the validation error curve to urge the optimum price of K. This price of K ought to be used for all predictions.

K-means clustering algorithm:

K-means is one in each of the sole unsupervised learning algorithms that solve the well-known cluster disadvantage. The procedure follows a simple and easy because of classify a given data set through an explicit style of clusters (assume k clusters) mounted apriori. The foremost set up is to stipulate k centres, one for each cluster. These centres have to be compelled to be placed in an exceedingly} very cunning suggests that because of utterly completely different completely different} location causes different result. So, the upper different is to place them the utmost quantity as potential isolated from each other. Succeeding step is to need each purpose happiness to a given data set and associate it to the nearest centre. Once no purpose is unfinished, the first step is completed associated Associate in

Nursing early cluster age is finished. At this point we'd prefer to re-calculate k new centroids as barycentre of the clusters succeeding from the previous step. Once we have these k new centroids, a novel binding must be done between identical data set points and so the closest new centre. A loop has been generated. As a results of this loop we have a tendency to tend to may notice that the k centres change their location step by step until currently tons of changes unit done or in several words centres do not move from now on.

B. SHA-256

SHA-256 is one in each of the successor hash functions to SHA-1 (collectively mentioned as SHA-2), and is one in each of the strongest hash functions getable. SHA-256 is not way more sophisticated to code than SHA-1, and has not all the same been compromised in any approach. The 256-bit key makes it Associate in Nursing honest partner-function for AES. it's printed inside the govt. agency (National Institute of Standards and Technology) traditional

C. RSA

RSA secret writing is usually utilized together with completely different secret writing schemes, or for digital signatures which could prove the legitimacy and integrity of a message. It isn't generally accustomed inscribe entire messages or files, as a results of it's less economical and tons of resource-heavy than symmetric-key secret writing.

To make things tons of economical, a file will generally be encrypted with a symmetric-key formula, then the regular key square measure encrypted with RSA secret writing. below this technique, alone Associate in Nursing entity that has access to the RSA personal key square measure ready to rewrite the regular key.

Without having the flexibility to access the regular key, the initial file can't be decrypted. this method is accustomed keep messages and files secure, whereas not taking too long or intense too many procedure resources.

5. Modules

A. User Registration and Authentication

In this module user needs to registers into his application and a request will be sent to central board server for authentication. Unless the central board server approves the request user cannot login into his account. When central board server approves the request a key will be generated and user can login into his account.

B. User Upload Certificate

After user login into his account he needs to upload certificates namely pan card, aadhar card, voter id, ssc certificates to central board server. Central board server will review the certificates and accepts or decline the certificates. If central board server accepts the accepts the certificate those details will be stored in E.C.S and Blockchain. If central board server declines the certificate it won't be stored in E.C.S. or Block Chain.

C. Get Certificate

If user needs a certificate he will send request to central board server. If central board server found the user details to be genuine he accepts the request and forward a request to E.C.S where all the certificates will be there. E.C.S. responds for the request and certificates will be provided to the user.

D. QR Request and Face verification

If user wants to apply for any certificates he will send request to central board server and central board server will check the details and forward the request to E.C.S. E.C.S will generate the QR Code and forwarded to user via central board server. User forwards the QR code to the verifying authority and if all details are correct and face matches with live face Verifying authority will issue the document.

6. Conclusion

In this paper, we tend to tend to propose a novel face recognition system named DocFace+, for cross- substantiating ID document photos to selfies. Here a certificate based system is developed supported blockchain technology to beat the issues of various selfie poses that square measure utterly completely different from general face recognition. By exploitation this technology anyone United Nations agency wishes to vary a particular internal information ought to together request for its various nodes at a similar time, since the information square measure hold on in many nodes. SHA256 with RSA signature is a cheap uneven secret writing methodology that initial calculates a singular hash of the digital documents exploitation SHA256 rule. The hash is then encrypted with a personal key exploitation the RSA rule. The system prevents document forgery, saves paper, cuts management costs, provides correct and reliable data on digital certificates and compare the user live face with verified document face.

References

[1] D. White, R. I. Kemp, R. Jenkins, M. Matheson, and A. M. Burton, "Passport officers' errors in face matching," PLoS ONE, vol. 9, no. 8, 2014.
[2] U.S. Customs and Border Protection. (2018). automatic Passport management (APC).
Available: https://www.cbp.gov/travel/uscitizens/apc
[3] Wikipedia. (2018). Epassport Gates.
[Online]. Available: https://en.wikipedia.org/wiki/EPassport_gates
[4] Wikipedia. (2018). Australia Smartgate.
[Online]. Available: https://en.wikipedia.org/wiki/SmartGate
[5] province HengAn Perimeter Security Instrumentation Company. (2018). what is ID-Person Matching?
[Online] Available: http://www.xjhazj.com/xjhazj/vip_doc/8380983.html
[6] V. V. Starovoitov, D. I. Samal, and D. V. Briliuk, "Three approaches for face recognition," in Proc. Int. Conf. Pattern Recognit. Image Anal., 2002, pp. 707–711.
[7] T. Bourlai, A. Ross, and A. K. Jain, "Restoring degraded face images: A case study in matching faxed, printed, and scanned photos," IEEE Trans. Inf. Forensics Security, vol. 6, no. 2, pp. 371–384, Jun. 2011.
[8] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces at intervals the wild: A data for learning face recognition in free

environments," Univ. Massachusetts, Amherst, MA, USA, Rep. 07-49, Oct. 2007.
[9] V. Starovoitov, D. Samal, and B. Sankur, "Matching of faces in camera footage and document pictures," in Proc. ICASSP, 2000, pp. 2349–2352.
[10] T. Bourlai, A. Ross, and A. Jain, "On matching digital face footage against scanned passport photos," in Proc. IEEE Int. Conf. Biometrics Identity Security (BIDS), 2009, pp. 1–10.
[11] Mitek. (2018). Mitek ID Verification. [Online]. Available: https://www.miteksystems.com/mobile-verify
[12] Jumio. (2018). Netverify ID Verification. [Online]. Available: https://www.jumio.com/trusted-identity/netverify
[13] Y. Shi and A. K. Jain, "DocFace: Matching ID document photos to selfies," in Proc. BTAS, 2018, pp. 1–8.