iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** II **Month of publication:** February 2025

DOI: https://doi.org/10.22214/ijraset.2025.67090

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue II Feb 2025- Available at www.ijraset.com

ScreenSafe: An AI-Powered KYC Name Screening Solution Using NLP

Yogeeta Suwasiya[1], Dev Patel[2]
Bachelor of Science in Data Science & Data Analytics, Vivekanand Education Soceity's College Of Arts, Science and Commerce (Autonomous), Mumbai

***Abstract:** This research paper is on ScreenSafe, a name screening tool for the KYC which uses NLP and other tools and apps for the OpenSanctions API for crooks' detection in finance industry and related areas. ScreenSafe uses NLP (Natural Language Processing) and the OpenSanctions API to automate name extraction and error detection with fuzzy matching and to have the bank issue a report for PEPs, thereby addressing the human inefficiencies in the manual process. This paper explains the design of ScreenSafe, the methodology employed, and the benefits gained from it in the KYC process.*
***Keywords:** KYC (Know Your Customer), natural language processing (NLP), named entity recognition (NER), anti-money laundering (AML), AI in financial services, and automated screening system.*

I. INTRODUCTION

The Name Screening[1][2] of KYC stands as the most important activity of the financial institutions to prevent money laundering, terrorist financing, and other illegal activities. KYC name screening[2][6] is a key operational risk reduction and security procedure used in financial institutions to detect and prevent illegal activities such as money laundering, terrorist financing, and the illegal drug trade. Traditional approaches involve manually cross-referencing customer names with global sanction lists, watchlists, and databases of politically exposed persons (PEPs). These traditional processes besides taking a lot of time are also prone to human error result in missing detection and wrong positive.

The applications which are built upon such advanced technologies are less likely to be obstructed by the regulatory authorities than the traditional manual methods. Yet, they can face major differences in the abilities of machine learning models against human interpretation skills that are not yet benefiting from advanced technologies.

The main result from the work of professional companies is that the risk mitigation cost is still high for the use of machine learning technologies even though many companies and the regulators have adopted this new technology. With an increase in the scrutiny of the regulation and the emerging of financial entities and threats, there is an urgent need for automatic and reliable systems that deal with a large amount of data and provide results accurately and quickly.

ScreenSafe is a tool that is intended for use by financial institutions that wish to benefit from Optical Character Recognition (OCR), Natural Language Processing (NLP), and solid API integrations. This tool is designed to make the KYC process easier by the automation of name extraction from various file formats, by getting rid of the probability of errors through the application of advanced fuzzy matching techniques, and by generating compliance reports which are comprehensive enough to satisfy the regulations.

II. OBJECTIVES

1) The key goal of the ScreenSafe platform is to fully automate the name extraction process[6] using the documents submitted by the customers. For example, this is accomplished by first using OCR to transform images or scanned documents into text and then using NLP methods to identify names. Automation in the name extract process simplifies the manual monotonous and error-prone job.

2) While automation has come a long way, there are cases where human intervention is absolutely required for the sake of accuracy. ScreenSafe lets users manually verify and modify the names fetched by the system. This step certifies that any abnormalities or mistakes with the automatic solutions are dealt with properly hence the integrity of the screening procedure is preserved.

3) Safety and precision in name indexing are what grabs the most attention. ScreenSafe connects to the OpenSanctions API[7], which is a universal source of lists of the sanctions and watchlists as well as PEP databases globally. This functionality ensures that all the name entries received have been checked and approved for any possible risk, thus the firm is in full compliance with AML[1][3] sanctions and FATF[4] regulations.

©IJRASET: All Rights are Reserved

4) It is important that the audit logs that are produced are fully detailed and are also downloadable for audit trails and regulatory filings. The ScreenSafe platform streamlines comprehensive reports to create visualizations that sum up the screening results, including the occurrence of matches. The reports also mention the actions taken and the overall compliance status. The internal audits and the compliance inspection to the regulatory officials are facilitated through these reports.

III. METHODOLOGY AND ALGORITHM

A. *Data Collection and Pre-processing*

1) Among other features, ScreenSafe can be used to convert documents of various types like PDFs, images (JPEG, PNG), and plain text files into machine-readable data. Also, this capability is the proof that the system can handle these kinds of files seamlessly.
2) The OCR process that is based on Tesseract, an optical character recognition system that is capable of recognizing texts from images, is applied for the conversion task. A document file which can't be edited and only seen on the computer screen can still be useful due to this way of converting the text.
3) In combination with the already known NLP library, spaCy, it identifies the Named Entity Recognition (NER) and explains how names are sought and retrieved from the given text. Pre-trained models by spaCy are adjusted for high accuracy in performing the task to recognize names under different contexts.

B. *Name Screening Workflow*

1) The names are fetched from the system and the extraction process is used to release the names that are required to be screened for the OpenSanctions API[7]. This API cross-checks extensive databases, such as sanction lists, watchlists, and PEP lists to locate potential matches. As part of the integration the most up-to-date and complete information is utilized for risk assessment.
2) In response to the problem of false positives as well as different variations of name spellings, the fuzzy matching algorithms are the ones that are involved. The algorithms compare names that are slightly different and give a similarity score which is helpful to reduce the number of false positives. [8]
3) A user-pleasant and eye-catching graphical user interface (GUI) is monitored to assist the user to check filter in clear feedback. Users will be able to keep track of the progress, check matches, and make relevant corrections that would help in the whole user experience increasing in the process.

C. *Project Analysis*

1) The accuracy of ScreenSafe is increased which is not common in standard NLP pipelines since they are not prepared for such a peculiar use-case. Custom NLP pipelines create the personalized name screening process that adapts to the variations of this domain and thus provides a higher level of accuracy. These pipelines are dynamically changed and made better as per the new data and scenario.
2) ScreenSafe is considered as the fastest name checking application because it can take only a few seconds to screen and process thousands of names that is what makes it faster than alternative manual approaches. This "screeners" which are time-efficient not only save the time spent but also allow us to deal with a considerable amount of data without delays.
3) By following the AML[1][3] and FATF[4] regulations, ScreenSafe guarantees that the screening process will be carried out according to the regulatory standards. Financial institutions are risk-free if they comply with this regulation because they will avoid paying unnecessary penalties, as well as be credible with government regulators.

IV. FINAL RESULTS

ScreenSafe sensibly assisting the separation of names, checking, and finally, reporting the result, thus improving KYC processes thereby reduces the likely risk of human error. The software tool's collaboration with OpenSanctions API[7] and its use of deep NLP technology[5] can show positive increase in both efficacy and propriety.

V. FUTURE ENHANCEMENTS

Migration of Django backend using Python for the development of new features for ScreenSafe, which include scalability, and robustness. The secure and scalable framework of Django will, in its support to ScreenSafe growth handles the capacity, even further provided, and more sophisticated systems will be installed.

©IJRASET: All Rights are Reserved

Extension of language support to allow for a broad array of users and cases to use this software. Through this feature, the application will be accessible to financial institutions in several different parts of the world and will bs serviced in several different languages.

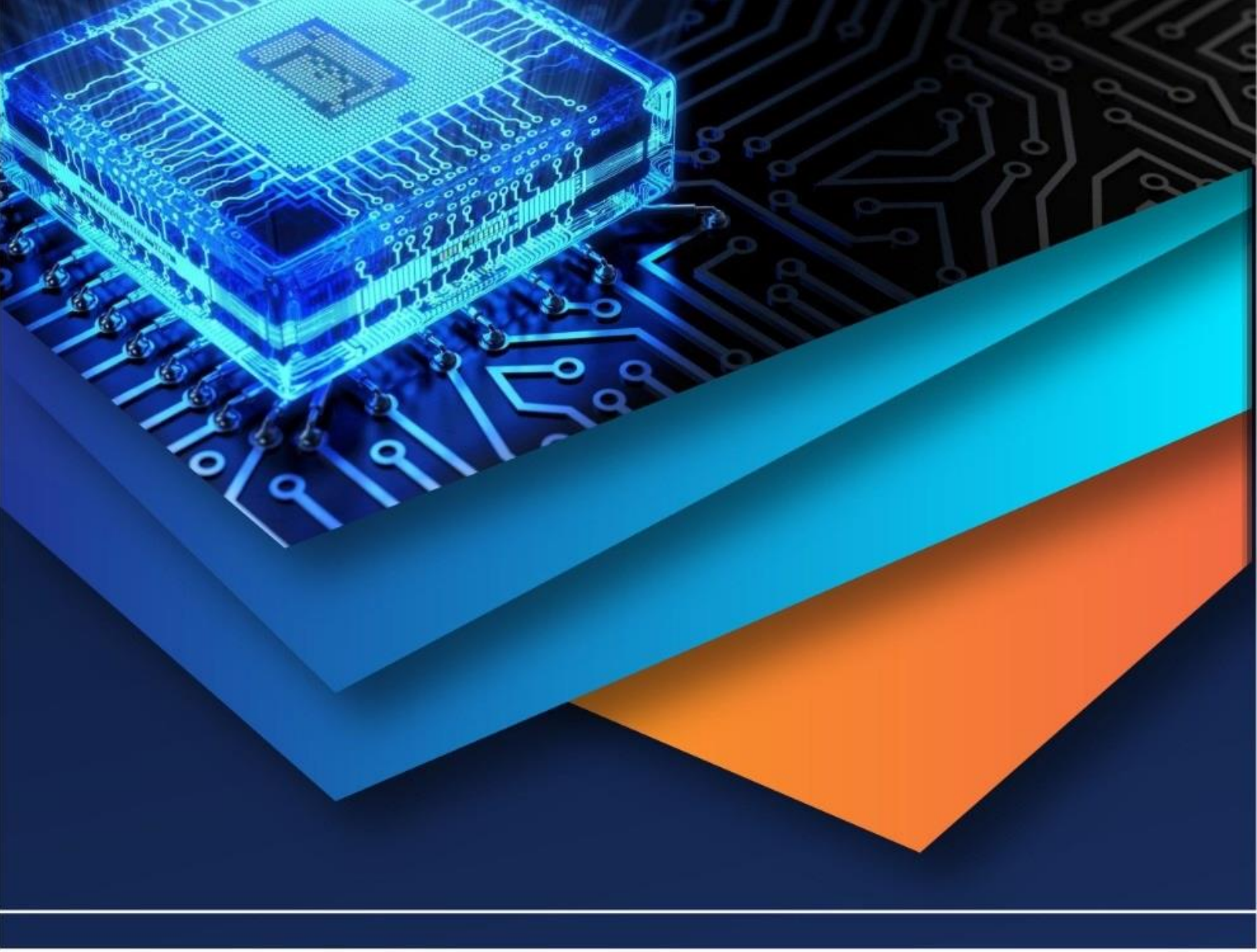
Role-based access control introductions are the key to reliable and controlled operations. This functionality will empower the different levels of access for a user based on his/her responsibilities, hence the purpose of security and compliance will be achieved.

VI. CONCLUSION

ScreenSafe has been found to be a significant breakthrough in KYC name screening with the advantages of newest NLP and API technologies. Automation, increased accuracy, and extensive reporting through ScreenSafe technology lighten the burden on financial service firms, woo and wax clients, and do much for preventing financial crime. ScreenSafe and tools alike have become indispensable for financial institutions to stay in compliance against the ever-changing threats and the increasing regulatory complexities.

REFERENCES

[1] Saperion, E. (2019). The Evolution of Name Screening in AML Compliance. Journal of Financial Regulation, 19(1), 45-60.
[2] Refinitiv. (2022). Know Your Customer (KYC) Screening and Monitoring. World-Check. Retrieved from KYC Screening
[3] Smartkyc. (2025). smartKYC's Response to MAS Guidelines - AML Name Screening Practices. Retrieved from https://smartkyc.com/smartkycs-response-to-mas-guidelines-on-strengthening-aml-cft-name-screening-practice/
[4] Financial Action Task Force (FATF). (2013). Guidance on Politically Exposed Persons (PEPs). FATF. Retrieved from FATF website
[5] MIT Sloan. (2025). Why finance is deploying natural language processing. Retrieved from https://mitsloan.mit.edu/ideas-made-to-matter/why-finance-deploying-natural-language-processing
[6] Smatbot. (2025). A Detailed Case Study on Automating KYC Document Verification Process. Retrieved from https://www.smatbot.com/blog/a-detailed-case-study-on-automating-kyc-document-verification-process/
[7] OpenSanctions. (2023). API Documentation and Implementation Guide. OpenSanctions. Retrieved from OpenSanctions website
[8] Sanctions.io. (2025). The Problem of Name Matching in Sanctions Screening. Retrieved from https://www.sanctions.io/blog/the-problem-of-name-matching-in-sanctions-screening



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)