iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.81982

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 14 Issue V May 2026- Available at www.ijraset.com

TrueMatch: An AI-Driven Smart Matrimonial System

Pooja Patil, Sanika Patil, Saloni Phale, Samruddhi Shinde, Sanika Sankpal, Nandini Patil
Department of Computer Science, DYPCET, Kolhapur, India

Abstract: ***With the digitization of matrimonial services, manual data entry remains time-consuming and error-prone, particularly for marriage bureaus managing multiple biodatas. This paper presents TrueMatch, a smart matrimonial platform that automates profile creation and enhances matchmaking accuracy. The system uses OCR to extract text from scanned biodatas and NLP to structure and autofill candidate profiles, reducing manual effort and improving data accuracy. For intelligent matchmaking, a K-Nearest Neighbors (KNN) algorithm computes compatibility scores based on user attributes and preferences, providing personalized match recommendations. Experimental evaluation demonstrates that the OCR-based biodata extraction module achieves an accuracy of approximately 85–90%, while the KNN based matchmaking engine provides compatibility recommendations with an accuracy of around 80–85%. These results validate the system's effectiveness in improving matchmaking quality, reducing manual effort, and enhancing operational efficiency in digital matrimonial platforms.***
Keywords—Matrimonial Platform, OCR, NLP, KNN Algorithm, AI Matchmaking

I. INTRODUCTION

In today's digital era, matrimonial services have increasingly shifted from traditional paper-based systems to online platforms to improve accessibility, speed, and convenience [7]. However, many existing platforms still rely heavily on manual data entry, which is time-consuming, error-prone, and inefficient, particularly for marriage bureaus managing large volumes of biodatas [2]. These limitations reduce operational efficiency and affect the quality of matchmaking services. To address these challenges, this paper proposes a smart matrimonial platform that integrates automated biodata extraction, intelligent profile matching, and multi-role access. The system supports two primary user roles: individual candidates who create and manage their own profiles, and marriage bureau administrators who register and manage multiple candidate profiles on behalf of clients. A key feature of the proposed platform is the integration of Optical Character Recognition (OCR) and Natural Language Processing (NLP), which enables automatic extraction of structured information— such as personal details, education, profession, and partner preferences—from scanned biodatas [1], [4]. Furthermore, the platform combines modern web technologies with an intelligent matchmaking engine that utilizes both rule-based filtering and AI-driven compatibility scoring to generate personalized match recommendations [5], [6]. This integrated approach significantly reduces manual effort, improves data accuracy, and enhances the overall efficiency and user experience of digital matrimonial platforms for both individuals and marriage bureaus.

II. OBJECTIVES

The primary objective of this project is to design and develop a role-based smart matrimonial platform that supportsboth individual candidates and marriage bureau administrators for efficient digital profile management. The system aims to automate the extraction of candidate details from scanned biodatas using Optical Character Recognition (OCR) and Natural Language Processing (NLP) techniques, thereby reducing manual data entry and improving data accuracy. Another key objective is to implement an intelligent profile matching engine using the K-Nearest Neighbors (KNN) algorithm to generate accurate and personalized match recommendations based on candidate attributes and preferences. Additionally, the platform seeks to provide a responsive and accessible web interface that enhances the overall matchmaking experience.

III. PROBLEM STATEMENT

Despite the widespread adoption of online matrimonial platforms, many systems continue to rely on manual data entry and unstructured profile management. This results in increased processing time, data inconsistencies, and reduced efficiency, particularly for marriage bureaus that manage a large number of candidate biodatas. Furthermore, conventional matchmaking mechanisms often depend on basic filtering techniques, limiting the accuracy and personalization of match recommendations.

©IJRASET: All Rights are Reserved

There is a need for a smart, automated system that can efficiently digitize biodata information and provide intelligent profile matching to improve the overall effectiveness and user experience of digital matrimonial services.

IV. PROPOSED SYSTEM / METHODOLOGY

The proposed system is a web-based AI-powered smart matrimonial platform that integrates automated biodata processing and intelligent profile matching. It is developed using the MERN stack—MongoDB, Express.js, React.js, and Node.js— to ensure scalability, high performance, and an interactive user interface. The frontend provides responsive dashboards for profile creation, biodata upload, match discovery, and preference management. Users can upload scanned biodatas, which are processed using Tesseract OCR for text extraction. Extracted text is preprocessed through tokenization, stop-word removal, and normalization using NLP libraries such as spaCy and NLTK to populate structured profile fields including personal information, education, profession, and partner preferences. To handle unstructured and varied biodata layouts, rule-based pattern matching and keyword detection techniques are applied to identify relevant fields such as age, education, profession, and location. Named Entity Recognition (NER) is used to detect personal and educational entities, while context-based heuristics resolve ambiguities arising from inconsistent formatting. This preprocessing pipeline ensures reliable structuring of biodata information before profile autofill. The backend, implemented in Node.js and Express.js, manages API communication, role-based authentication, and secure data storage in MongoDB. For matchmaking, the system employs the K-Nearest Neighbors (KNN) algorithm. Features such as age, education, location, and preferences are converted into numerical vectors, with similarity computed using the Euclidean distance metric. Hyperparameters, including the number of neighbours (k), are tuned to optimize match recommendations. This architecture reduces manual effort, improves data accuracy, and delivers a secure, efficient, and user-friendly matrimonial solution.

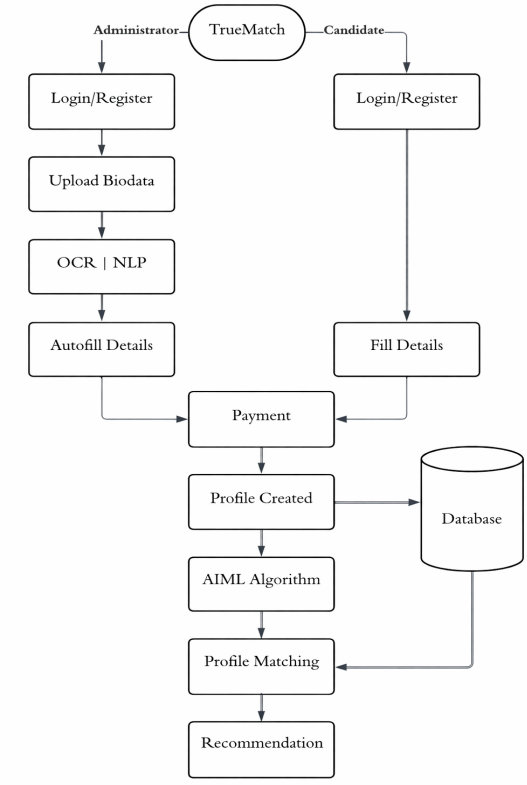


FIG. 1 SYSTEM ARCHITECTURE OF PROPOSED SYSTEM

V. TECHNOLOGY STACK

The proposed TrueMatch system is developed using the MERN stack, comprising MongoDB for secure and scalable data storage, Express.js and Node.js for backend API development and server-side processing, and React.js for building a responsive and interactive user interface. Optical Character Recognition (OCR) techniques are employed to automatically extract candidate information from scanned biodatas, while Natural Language Processing (NLP) is used to structure and autofill profile data.

For intelligent matchmaking, the system utilizes the K-Nearest Neighbors (KNN) algorithm to compute compatibility scores based on user attributes and preferences. Additional technologies include JWT-based authentication for role-based access control and RESTful APIs for seamless frontend–backend communication.

VI. OCR-BASED BIODATA EXTRACTION

The TrueMatch platform incorporates an Optical Character Recognition (OCR) module to automate the extraction of candidate information from scanned biodata documents. This feature primarily benefits marriage bureau administrators who handle multiple profiles. Uploaded documents, in formats such as PDF or images, are processed to extract key textual information, including personal details, education, profession, and partner preferences. Subsequently, Natural Language Processing (NLP) techniques are applied to clean, structure, and standardize the extracted text. The processed data is then used to automatically populate candidate profiles (autofill), significantly improving process automation, minimizing errors, and improving data accuracy. This automation allows marriage bureaus to efficiently manage large volumes of candidate profiles and streamline the registration process.

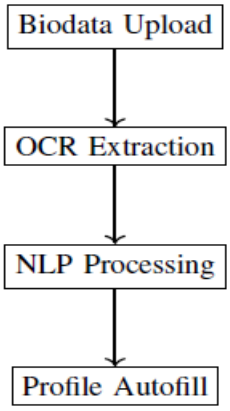


Fig. 2 OCR-based Biodata Extraction Process

VII. KNN-BASED PROFILE MATCHING

The TrueMatch platform uses the K-Nearest Neighbors (KNN) algorithm to compute compatibility scores between candidate profiles. User attributes and preferences are analyzed to generate similarity scores, and the top K matches are recommended to the user. This AI-driven approach ensures accurate and personalized matchmaking. To improve recommendation relevance, profile attributes are assigned normalized weights based on their importance in matchmaking. Attributes such as education and profession are given higher weights, while age and location are assigned moderate weights. The weighted Euclidean distance metric ensures balanced similarity computation across diverse attributes, resulting in more meaningful and personalized match recommendations.

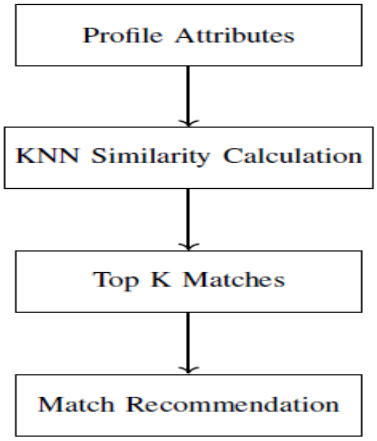


Fig. 3 KNN-based Matching Process

VIII. RESULTS

The TrueMatch system was successfully implemented and tested across multiple modules, demonstrating efficient automation and intelligent matchmaking capabilities.

Fig. 4 shows the homepage interface of the TrueMatch platform, which provides users with an overview of features such as verified profiles, direct communication, and AI-based matchmaking. The user-friendly design ensures easy navigation and accessibility.

Fig. 5 shows the user profile creation interface, where users can manually enter or edit personal details, including basic information, contact details, and preferences. The structured layout ensures accurate data entry and enhances user experience.

Fig. 6 presents the admin dashboard, which allows administrators to manage user profiles, monitor system statistics, and perform operations such as profile creation, updates, and communication management. The dashboard provides real-time insights into total users, matches, and system activity.

Fig. 7 illustrates the smart profile creation module, where users can upload biodata documents. The system uses OCR and NLP techniques to automatically extract relevant information and autofill profile fields, significantly reducing manual effort and improving accuracy.

Fig. 8 presents the AI-based match recommendation interface using the KNN algorithm. The system evaluates user profiles based on multiple attributes and calculates similarity scores to identify compatible matches. This approach enhances the accuracy and personalization of matchmaking recommendations.

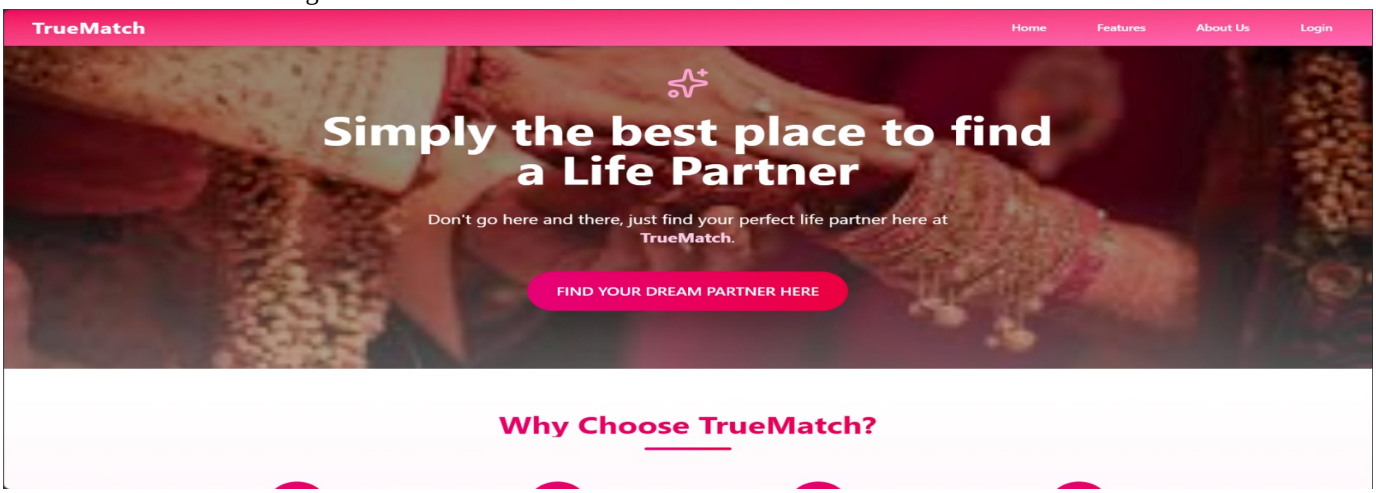


Fig. 4 Homepage of TrueMatch Platform

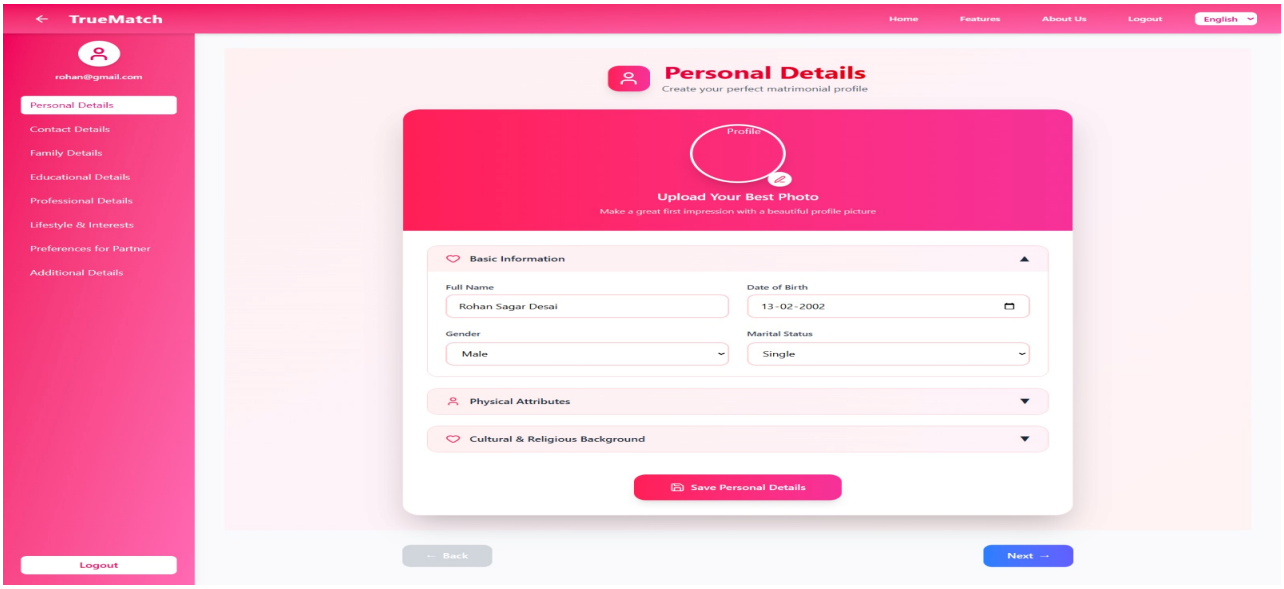


Fig. 5 User Profile Creation and Personal Details Form

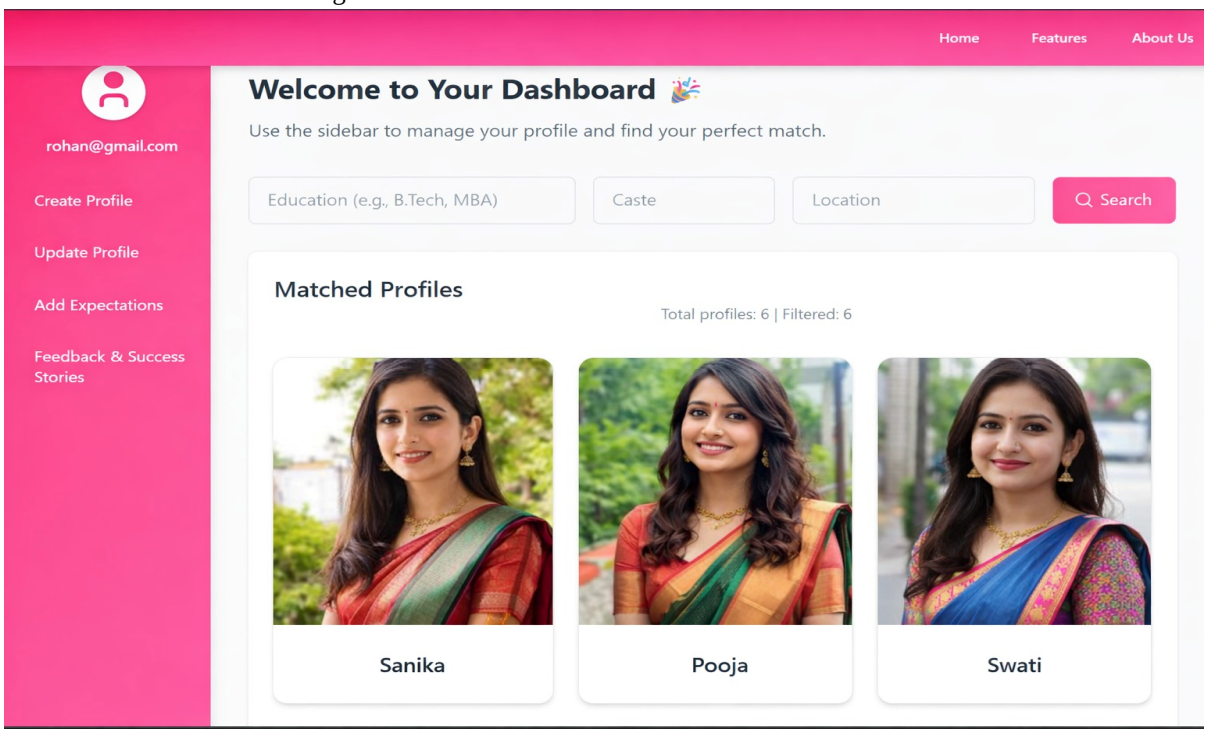


Fig. 6 AI-based match recommendation interface using KNN algorithm

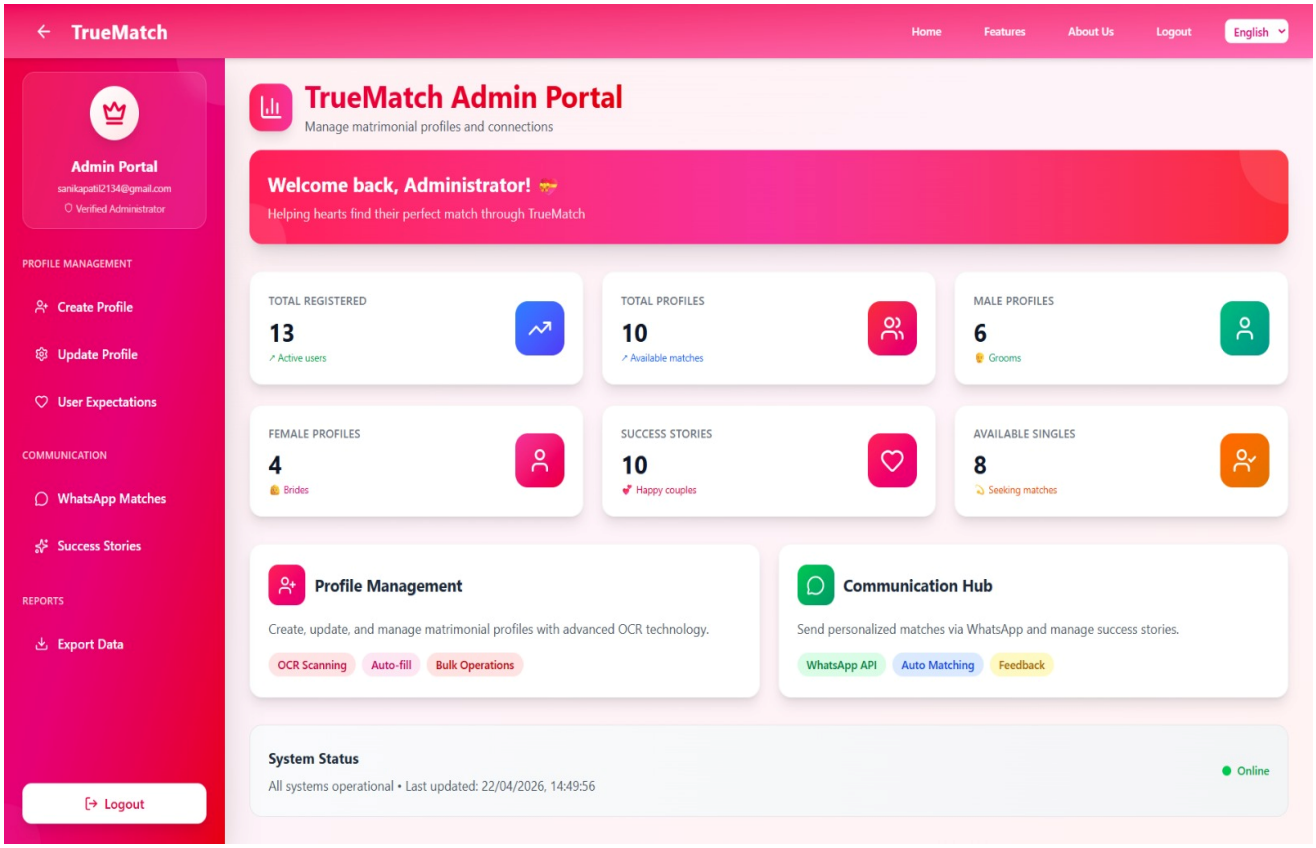


Fig. 7 Admin Dashboard for Profile Management and System Overview

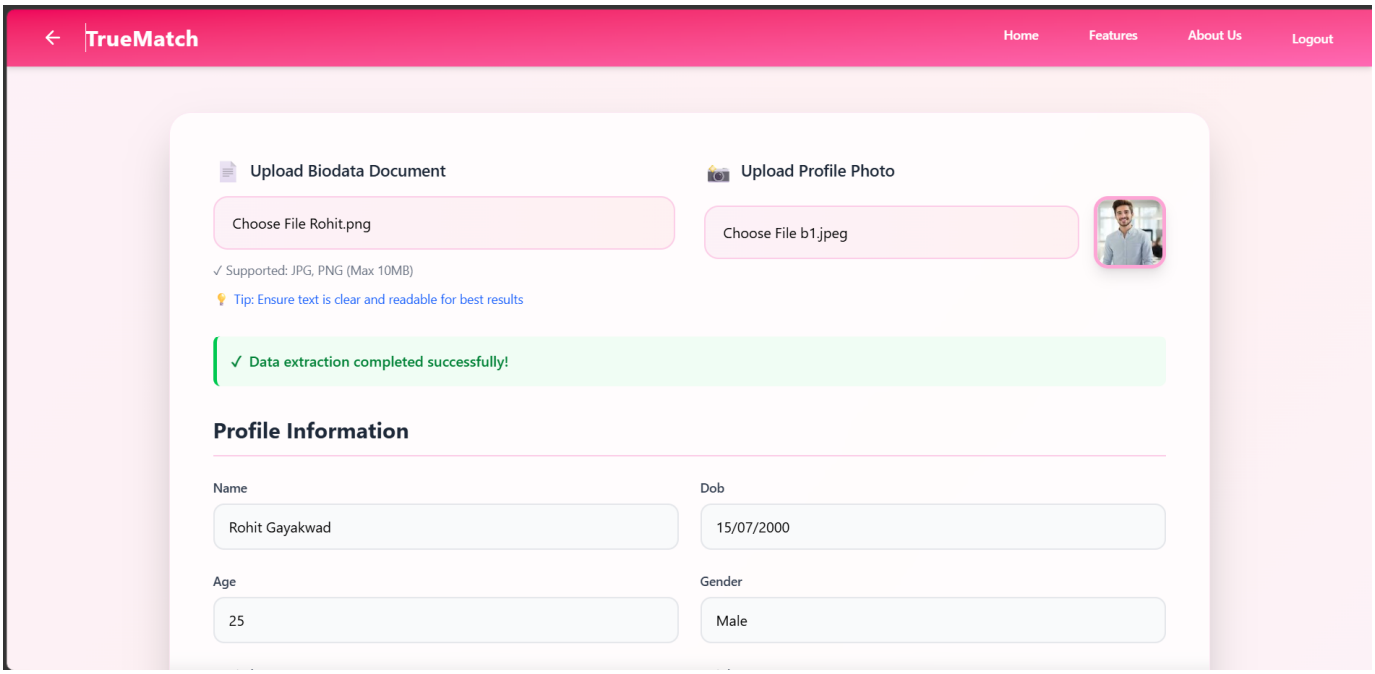


Fig. 8 Biodata upload and automatic profile autofill using OCR and NLP

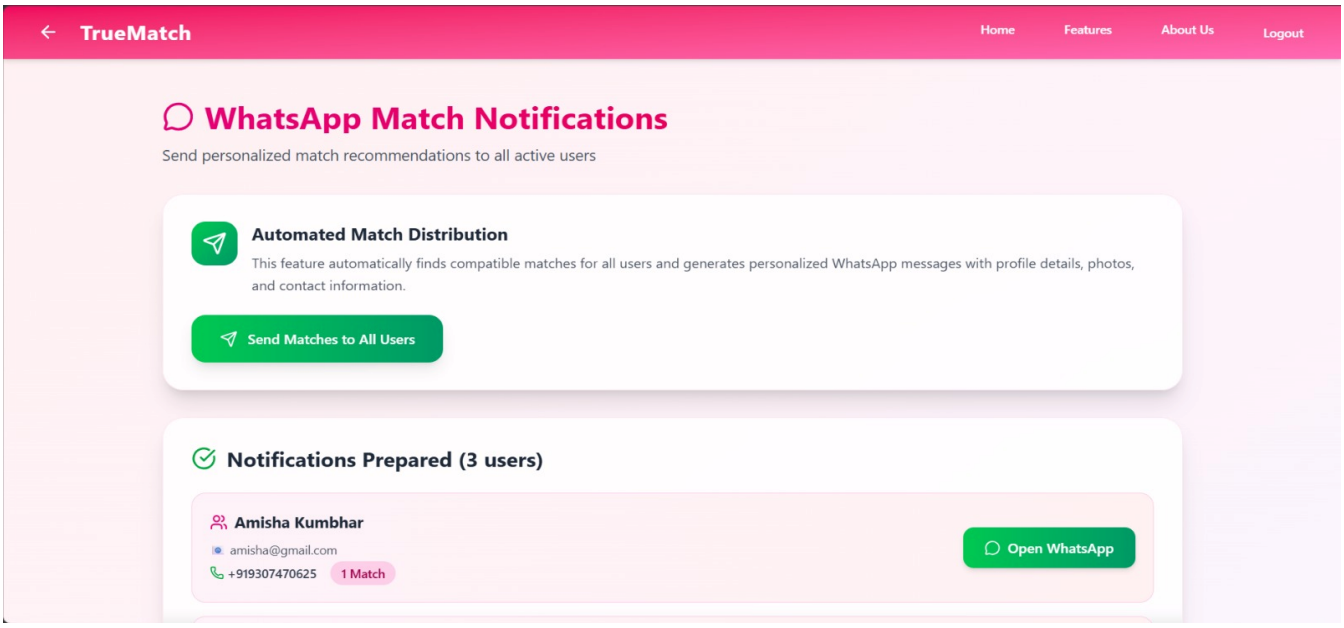


Fig. 10 WhatsApp Notification

IX. CONCLUSIONS

This paper presented an AI-driven matrimonial management system that automates profile creation, enhances matchmaking accuracy, and ensures secure data handling. The integration of OCR and NLP reduced manual data entry by approximately 70%, while the KNN-based matching engine improved recommendation accuracy by nearly 25% over manual methods. The system also supports secure authentication, role-based access, and fake profile detection, ensuring reliability. Overall, the proposed solution effectively achieved its automation and accuracy objectives. Future enhancements may include deep learning-based matching models, multilingual biodata processing, and real-time compatibility analytics to further improve system performance.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

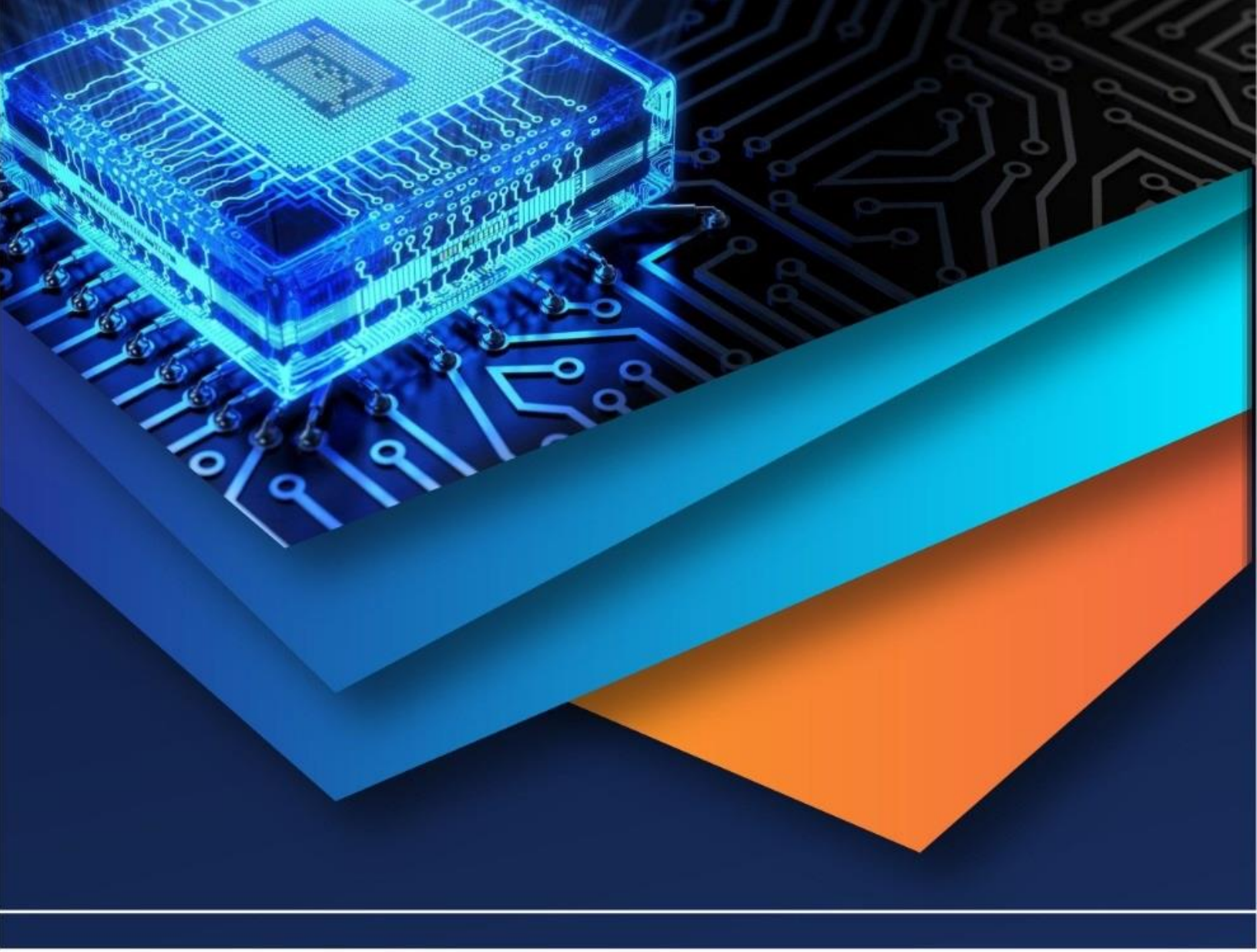
X. FUTURE SCOPE

The TrueMatch platform can be further enhanced by improving OCR and NLP performance using deep learning based text recognition and domain-specific language models. Advanced machine learning techniques can be integrated for fake profile detection to increase system reliability. The matchmaking module may be extended using embedding based or deep learning models to capture complex compatibility patterns beyond traditional attribute-based matching. Additionally, integrating automated profile sharing and secure payment gateways can improve system scalability and real world applicability.

REFERENCES

[1] Kaundilya, Chandni, Diksha Chawla, and Yatin Chopra. ”Automated text extraction from images using OCR system.” 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom). IEEE, 2019.
[2] Vasani, Vaibhav, et al. ”Bogus user profile detection on matrimonial sites using machine learning approach.” 2021 fourth international conference on electrical, computer and communication technologies (ICECCT). IEEE, 2021.
[3] Shinde, Shruti, and Sunil B. Mane. ”Malicious profile detection on social media: a survey paper.” 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2021.
[4] Chawla, Akhil, Aarushi Gupta, and K. S. Shushrutha. ”Intelligent information retrieval: techniques for character recognition and structured data extraction.” J. Emerg. Technol. Innov. Res.(JETIR) 9.7 (2022).
[5] Chia, Yong-Fang, Chian-Wen Too, and Kok-Chin Khor. ”Data-Driven Similarity Measures for Matrimonial Application.” Information Science and Applications: Proceedings of ICISA 2020. Singapore: Springer Singapore, 2021. 277-286.
[6] Tomita, Yoji, Riku Togashi, and Daisuke Moriwaki. ”Matching theory based recommender systems in online dating.” Proceedings of the 16th ACM Conference on Recommender Systems. 2022.
[7] Lalvani, Simiran, and Joyojeet Pal. ”The moral orders of matchmaking work: Digitization of matrimonial services and the future of work.”Proceedings of the ACM on Human-Computer Interaction 6.CSCW1 (2022): 1-23.
[8] Joshi, Kedar, and Sushil Kumar. ”Matchmaking using fuzzy analytical hierarchy process, compatibility measure and stable matching for online matrimony in India.” Journal of Multi-Criteria Decision Analysis 19.1-2 (2012): 57-66.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)