AI based Match Maker using Recommendation System and K-Means Clustering Algorithm

[1]Arshiya Fathima, [2]Steve Prince, [3]Stalin Varghese K, [4]Syed Affan,[5]Tenzin Choyang[5]
[1]Assistant Professor, [2,3,4,5]*UG Student,* [1,2,3,4,5]*Computer Science and Engineering*
HKBK College Of Engineering Bengaluru, India.

Abstract—This project presents an AI-driven matchmaking system that utilizes supervised and unsupervised learning techniques alongside Convolutional Neural Networks (CNN) for facial recognition-based authentication. The primary goal is to facilitate personalized matchmaking for marriage purposes through comprehensive profile analysis and user interaction. Supervised learning algorithms categorize user-provided information spanning Movies, TV, Religion, Music, Sports, Books, and Politics, while Natural Language Processing (NLP) techniques extract meaningful insights from user bios. Unsupervised learning methods uncover latent patterns within the dataset, enhancing matchmaking accuracy. Additionally, CNN-based facial recognition ensures secure user authentication during login. The system creates comprehensive user profiles, analyzes data, generates compatibility scores, and ranks potential matches, presenting users with curated lists of top matches. Integrated chat functionalities facilitate communication between matched individuals, promoting active engagement. This innovative system represents a pioneering initiative in personalized relationship management, leveraging AI technologies to redefine social interaction and relationship-building in the digital age.

Index Terms—K-mean Clustering, CNN, AI, NLP, Match Making, supervised, unsupervised.

INTRODUCTION

Online Match-making is indeed one of the fastest-growing facets in social platforms. It involves a unique-form of recommendations. And In contrast, to recommendations that primarily suggest items to users online, matchmaking recommendations rely on the requirement of shared interests between the two users to facilitate communication, from the outset [3]. One of the significant/biggest challenges lies in efficiently identifying matches for a user among the millions of users present on the online matchmaking networks [4]. The most prevalent recommendation systems, specifically content-driven and collaborative-based systems, are widely used, but also do come with its limitations. In the realm of online match-making recommendations, there has been limited explorations. The authors of a study applied an existing collaborative recommendation technique from an online dating platform, where we can employ the user's rating data. However, crucial factors like age, education, occupation, ethnicity etc. which significantly impact matchmaking, were overlooked in this study, resulting in low accuracy [5]. The matchmaking platform also contributes and leads towards the growth, of the creative economy sector, anticipated to support the business landscape especially in Indonesia and the other emerging countries. This growth will lead to a rise in exports of products, promote job opportunities and at the same time support the growth of startup education whether its, in its early phases or already established [15]. To improvise user satisfaction, on platforms we can use recommender systems. They help overcome the limitations of search functions and matching based on user profiles. By organizing and prioritizing matches, these systems make it easier for users to find their match. This streamlined approach promotes their interactions and communication with matches benefiting users who may not be as prominent on the plat- form. Moreover, by suggesting connections involving these less visible users when suitable, there's potential to reduce the overwhelming volume of unwanted communication. It's important to note that such a recommender system wouldn't replace existing search tools, but rather complement them, serving as an augmentation to the current interface [2].

Content-based recommender systems, generate suggestions for users based on an analysis of their past item selections, forming a profile from this content. These methods, originating from studies in information retrieval and machine learning, utilize various tactics to extract and evaluate user profiles

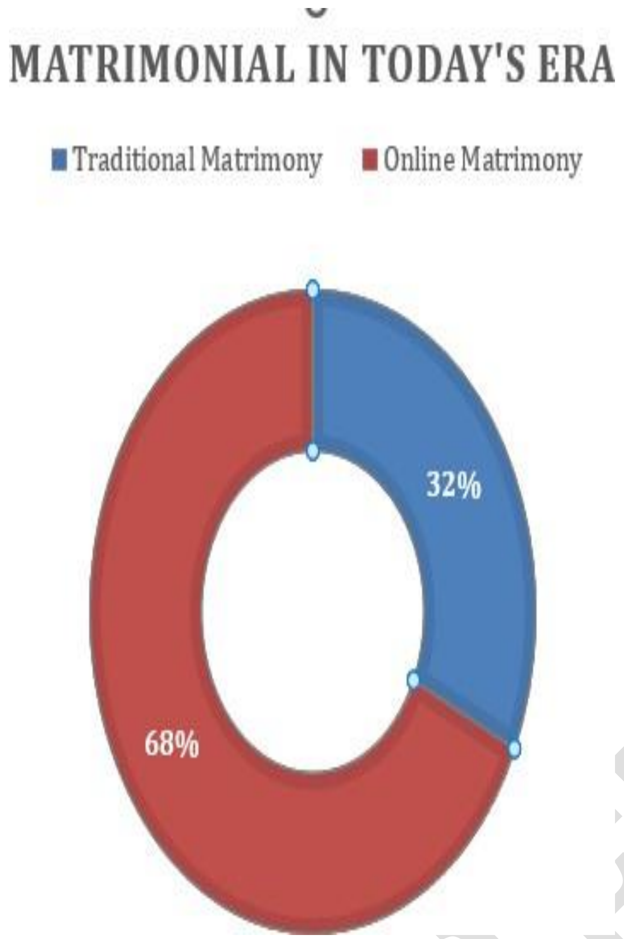


Fig. 1. Matchmaking Data

and item content. Nonetheless, early experiments uncovered difficulties in utilizing user profiles for content-based suggestions due to inadequate or disorganized information provided by users, particularly in free-form text, posing challenges for dependable analysis. Additionally, a significant drawback of content-based systems is their inability to recommend new items of potential interest to users, as they solely rely on suggesting items akin to a user's past preferences [2].

An important aspect of utilizing matchmaking platforms involves creating a matrimonial profile, typically involving the upload of photos and information which includes name, age, social media platforms etc., to create a positive image. Studies indicate that people highly prioritize their dating profiles, intending to make a positive impression on potential partners. The primary goal is often to transition online connections into real-life meetings. Nevertheless, building a matrimonial profile poses a lot of challenges and requires a significant and immense number of efforts, taking in the manual input of information while selecting content to attract potential matches. Additionally, study shows that individuals have a hard time creating appealing profiles as they move towards the delicate balance between showcasing positive traits and their authenticity to others [11]. Finding and stopping fake user accounts stay as a challenging task. In a social network that boasts millions of engaged users and countless user interactions, the number of false accounts is extremely low. In order to prevent unintentionally blocking genuine users it is imperative to maintain low rates of false positives. Certain fraudulent accounts exhibit indications of being automated. Many of them are deliberately created to mimic ones. Security measures like SMS-based phone verification and Captchas aim to investigate accounts that seem suspicious, thereby heightening the obstacles associated with creating fake profiles [6].

At the heart of this approach lies the incorporation of K-means clustering methodology. This technique provides a unique advantage by enabling the system to identify inherent patterns and groupings within the dataset. By categorizing individuals into distinct clusters based on shared traits, the system can offer a nuanced understanding of compatibility dynamics. The utilization of K-means clustering goes beyond conventional models, allowing for a more precise and tailored matchmaking experience. As individuals seeking life partners exhibit diverse preferences and characteristics, the inclusion of K-means clustering serves as a key mechanism to enhance the granularity and accuracy of the prediction model.

In addition to clustering, the system integrates Convolutional Neural Networks (CNNs) for facial recognition during sign-up and sign-in processes, ensuring secure and seam- less user authentication. Furthermore, classification algorithms such as K-Nearest Neighbors, Support Vector Machines (SVM), and Random Forest are utilized for classifying new profiles and detecting fraudulent accounts, thereby enhancing the reliability and security of the matchmaking platform. Naive Bayes algorithm is employed for recommendation and classification tasks, providing an efficient and probabilistic approach to modeling user preferences and identifying potential matches.

Through assessing the precision, effectiveness, and ethical implications of the system, our objective is to guarantee the dependability and impartiality of the recommendations it offers. The Research not only addresses the contemporary challenges in traditional matchmaking but also explores the potential to foster more harmonious and enduring relationships. As we navigate the intersection of technology and human relation- ships, this project represents an effort towards creating a more informed and adaptive approach to marriage matchmaking, promising a future where compatibility predictions are not only accurate but also ethically sound and tailored to the individual.

This model mainly applies the K-means clustering approach. This methodology shows a unique advantage by enabling the system to identify inherent patterns and forming groups within the dataset. By categorizing people into clusters based on their characteristics and traits the system provides a deeper understanding of how compatibility works, the use of K- means clustering goes beyond methods enabling an accurate and personalized matchmaking experience. Since people have preferences and characteristics when looking for life partners, incorporating K-means clustering is a way to improve the precision and detail of the prediction model.

We analyze the system's precision, effectiveness, and ethical implications to guarantee the dependability and impartiality of the suggestions given. This study is not just looks into challenges in traditional matchmaking but also explores the potential to foster more harmonious and enduring relation- ships. In our journey through the convergence of technology and human connections, this undertaking signifies an endeavor to develop an informed and flexible method for matchmaking in marriages. it shows that predictions about compatibility are not just precise but ethical and customized to each user.

I. Related Works

In today's digital realm, personalized recommendation systems are crucial for enhancing user experiences across diverse online platforms. From social software to matchmaking net- works, these systems employ innovative techniques to deliver tailored suggestions, optimizing user engagement. This paper explores their evolution and significance, from social software recommendations to AI-driven matchmaking platforms. The research paper titled "Personalized recommendation of social software items based on social relations" introduces a system that utilizes information from networks to offer personalized recommendations. The researchers carried out an investigation involving 290 Lotus Connections users to evaluate familiarity and similarity networks as a basis for suggestions. Additionally, they examined the significance of explanations based on individuals and compared various categories of suggested items. The conclusion of this paper is that individuals found the recommendations beneficial, with positive ratings for interesting and relevant items [1]. The research paper titled "Interaction-Based Collaborative Filtering Methods for Recommendation in Online Dating" proposes collaborative filtering techniques to improve user satisfaction in online dating websites. Numerous novel techniques and metrics are proposed to gauge and forecast potential improvements in user interaction satisfaction. The study explores different collaborative filtering techniques, such as Basic Collaborative Filtering (Basic CF), Inverted Collaborative Filtering (Inverted CF+ Sender and Inverted CF+ Recipient), and a combination of these approaches like Combined CF+ and Best Two CF+. Rigorous experiments conducted on past data obtained from a commercial online dating platform reveal substantial enhancements in user success rates [2]. The study titled "Improving Matching Process in Social Networks" delves into improving social network matching through the utilization of the Sim- Rank algorithm for assessing user similarity. It discusses user recommendations derived from these evaluations, experimental findings, and other pertinent research in social network analysis and user engagement. Tests conducted on a dataset sourced from a dating platform boasting around 2 million registered users illustrate the efficacy of the proposed approaches[3]. The paper entitled "A Recommendation Method for Online Dating Networks Based on Social Relations and Demographic Information" presents an approach to create a recommendation system tailored for online dating platforms. It tackles the issue of effectively pairing users in such networks, dealing with extensive yet sparsely populated datasets and the necessity for bidirectional matching. By integrating social networking principles, clustering techniques, and employing Simrank and, modified SimRank algorithms, the system suggests potential matches. According to empirical results, this approach exhibits substantial enhancements in performance when contrasted with conventional methods[4]. The research paper titled "Reciprocal Recommendation System for Online Dating" introduces similarity features that capture the characteristics of dating networks. The proposed recommendation systems outperform existing methods, with improvements in precision and recall for both female users. The content-based algorithm CB2 shows enhancements in evaluating attractiveness and interest between users, while the hybrid collaborative filtering algorithm (HCF) demonstrates superior performance in suggesting users likely to engage in communication. Additionally, the content-based reciprocal algorithm

RECON offers valuable insights into user preferences and behaviors, outperforming other methods in providing reciprocal recommendations [5]. The study entitled ”Detecting Clusters of Fake Accounts in Online Social Networks” explores the identification of clusters of false profiles in online social networks(OSNs) through machine learning techniques. The authors introduce a scalable and time-sensitive machine learning method for detecting groups of false profiles created by the same entity. They employ a supervised machine learning pipeline comprising the Cluster Builder, the Profiler Featurizer, and the Account Scorer, achieving an AUC of 0. 98 on a withheld test set and 0. 95 on unseen data testing data using the random forest algorithm. This approach has been implemented in production and has identified over 250,000 false profiles since its deployment [6]. The research paper titled “A Survey of Collaborative Filtering-Based Recommender Systems from Traditional Methods to Hybrid Methods Based on Social Networks” delves into recommendation systems (RS) and their applications across various domains. It explores traditional collaborative filtering (CF) recommendation methods, including memory-based and model-based CF methods, as well as latent factor models (LFM) and its variations such as matrix factorization, non-negative matrix factorization , and singular value decomposition (SVD). The document also explores the diverse applications of recommender systems across various domains, encompassing news platforms, social media networks, video streaming platforms, and music streaming platforms[7]. In a research paper titled “Switching Strategy of Recommendation Algorithms in Online Dating Platforms,” an approach to transitioning between different recommendation algorithms based on user feedback and preferences is presented. The document presents concepts of Recommended User Algorithm (RUA) and Recommended Algorithm Algorithm (RAA), along with describing the system architecture and data sources for their application. It discusses five strategies for RAA and four algorithms for RUA. The effectiveness of the proposed approach is assessed using metrics like MAE, precision, recall, and coverage. The results indicate that the multi-algorithm switching approach attains both accuracy and coverage, high- lighting its flexibility in accommodating user preferences and circumstances [8]. The research paper titled “Sharing and Privacy in Dating Apps” explores sharing and privacy issues in online matchmaking applications. The document provides strategies for app developers to improve user experience while addressing issues of sharing and privacy. It encompasses a review of existing literature and an examination of popular dating platforms like Tinder and Okcupid to pinpoint sharing functionalities that enhance social presence, trustworthiness, and intimacy, alongside significant privacy issues. It delves into various forms of sharing, such as geolocation, photo sharing, personal descriptions, and interests, while also Investigating various privacy issues associated with disclosing results and third party tracking systems [9]. The research paper titled “Marriage Recommendation Algorithm Based on KD-KNN-LR Model.” proposes a method that combines KD KNN for user classification and LR for reverse classification to find suitable matches. Factors such as age, education, and location for male and female users are taken into account. Our algorithm achieves an accuracy of 86 percent in recommending candidates who are likely to be accepted by both parties. Further improvements are suggested by refining feature representation and incorporating model regularization techniques [10]. In a research study titled “Online Dating Meets Artificial Intelligence; How the Perception of Algorithmically Generated Profile Text Impacts Attractiveness and Trust,” an investigation was conducted to examine the influence of AI involvement in creating reliable profile descriptions. The results showed that, although the Understanding of AI participation didn’t notably impact perceived attractiveness, it did cause a decline in the perceived trustworthiness ascribed to the profile creators. Further analysis incorporating AI into other areas may provide insights into the relationship between AI and online dating [11]. The research study titled “Application of Machine Learning to Create a Recommendation in Social Communication Based on Data Analysis” proposes the use of neural networks to analyze various attributes like attractiveness, age, and gender from user photos on Tinder. Additionally, the study involves object identification within these photos to generate keywords, which can then be combined with confidence scores obtained from photo analysis to generate recommended first message texts [12]. The research work titled “Detecting Bogus User Profiles on Matrimonial Sites Using Machine Learning Techniques” evaluates the effectiveness of regression random forest, XGBoost, and SVM classifiers in identifying profiles on matrimonial platforms. The random forest achieves an accuracy rate of 93 percent, showing excellent precision in classifying test profiles compared to other models [13]. The research paper titled “Predicting romantic interest during early relationship development-A preregistered investigation using machine learning” captures the early relationship develop- ment phase, where individuals experience rising and falling romantic interest for potential partners. Random forests are used to estimate the predictors of participants’ romantic inter- est in these potential partners, revealing robust main effects for many variables, including perceptions of the partner’s positive attributes and perceived interest [14]. The paper ”Artificial Intelligence Based on Recommendation System for Startup Matchmaking Platform” delves into employing an AI-driven recommendation system to streamline startup matchmaking, with the goal of crafting a profoundly smart platform capable of aiding the industry in information retrieval through startup matchmaking. Additionally, it explores Agile Sprint methodologies in software development and evaluates startup matchmaking platforms utilizing artificial intelligence and machine learning [15]. This paper titled “Design of a Face Recognition System based on Convolutional Neural Network (CNN)” presents The creation and execution of a facial recognition system utilizing Convolutional Neural Networks (CNNs). The suggested CNN structure comprisesof two convolutional layers, a fully connected layer, and a classification layer utilizing the Softmax function. The system attains a training accuracy of 99.78% and a validation accuracy of 98.75 % with the ORL face dataset, surpassing various cutting-edge approaches. [16]. The paper titled “A bio-inspired application of natural language processing: A case study in extracting multiword expression” presents a bio-inspired approach to multiword expression (MWE) extraction using multiple sequence alignment (MSA). The MSA method has higher recall but lower precision than n gram based approaches like the positional n-gram model (PNM). The study shows that MSA outperforms PNM for longer MWEs over 4 words and incorporates linguistic knowledge through error-driven learning rules to improve extraction for infrequent MWEs [17]. The paper titled “Implementation of Recommendation Systems in Determining Learning Strategies Using the Na¨ıve Bayes Classifier Algorithm” presents the development and implementation of a recommendation system to determine effective learning strategies for students based on their learning preferences. Collaborative filtering techniques and the Na¨ıve Bayes algorithm are used to provide recommendations to teachers. The system achieves an accuracy of 90.91% in determining appropriate learning strategies aligned with student learning styles, contributing to the improvement of education quality [18].

II. PROPOSED WORK

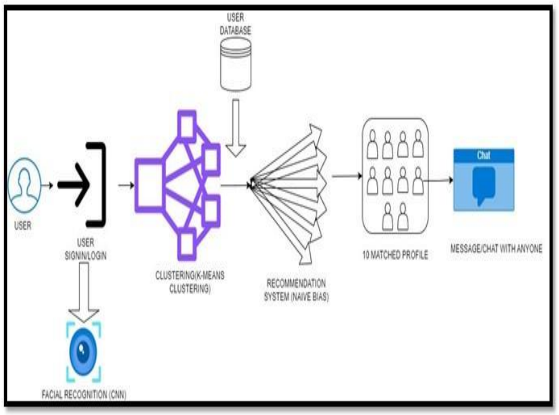


Fig. 2. Matchmaking Architecture

This methodology outlines a user authentication and recommendation system designed to enhance user experience and foster connections. Here’s a breakdown of each step:

User Authentication: The process begins with facial recognition using a Convolutional Neural Network (CNN). This advanced technology identifies the user based on their unique facial features, ensuring secure access to the system. User Sign-In/Log-In: Once the user is recognized, they proceed to sign in or log into the system, providing authentication for accessing personalized features and functionalities.

Data Clustering: User data is then clustered using K-Means clustering. This technique groups similar users together based on their attributes, such as interests, demographics, or behavior patterns.

User Database: The clustered data is stored in a user database, which contains profiles and relevant information about users. This structured database facilitates efficient data retrieval and management.

Natural Language Processing (NLP) for Bio Extraction: NLP techniques are employed to extract information from user bios. This involves parsing and analyzing textual data to identify key attributes, interests, and preferences mentioned in the user profiles.

Recommendation System: A recommendation system with a Naive Bayes classifier selects 10 matched profiles from the database. These profiles are recommended to the user based on shared interests, preferences, or other relevant factors identified through clustering analysis.

Interaction: Finally, the user can engage with anyone from the recommended profiles through messaging or chat functionalities. This interaction feature facilitates connections between users with shared interests or compatibility, fostering meaningful interactions and potential relationships.

A. User Authentication Using Convolutional Neural Networks (CNNs):

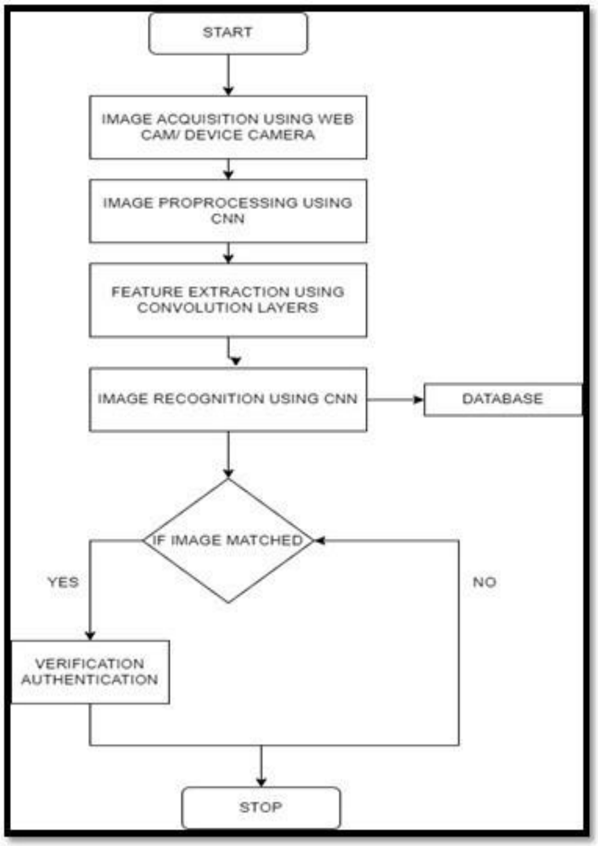

START
IMAGE ACQUISITION USING WEB CAM/ DEVICE CAMERA
IMAGE PROPROCESSING USING CNN
FEATURE EXTRACTION USING CONVOLUTION LAYERS
IMAGE RECOGNITION USING CNN
DATABASE
IF IMAGE MATCHED
YES
NO
VERIFICATION AUTHENTICATION
STOP

Fig. 3. CNN Flow Chart

The integration of CNNs for user authentication involves several steps. Initially, a diverse dataset of facial images is collected and annotated with labels. Images undergo preprocess- ing, including resizing and normalization. The architecture of CNN comprises input, convolutional, pooling, fully connected, and output layers. Feature learning through convolution and pooling operations enables automatic feature extraction. The classification stage involves flattening the features into a vector and passing them through fully connected layers for predic- tion. Training involves dataset splitting, model initialization, backpropagation, and validation. Data augmentation further improves model generalization. Face detection and alignment using Haar cascades facilitate accurate recognition. The image

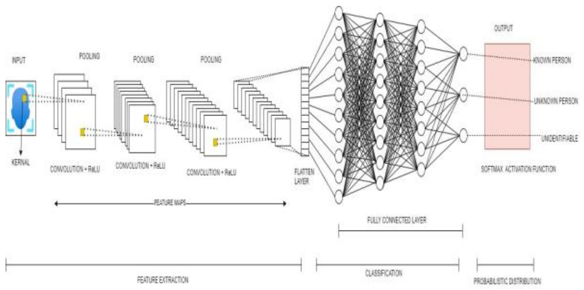


Fig. 4. CNN Architecture

shows a convolutional neural network (CNN) based architecture for image classification. Here's a breakdown of the general steps involved in this architecture:

Input Image: This is the image that the network is trying to classify. For instance, it might be an image containing a handwritten digit, or a picture of a cat.

Convolution Layer: The image undergoes a sequence of convolutional layers. Within these layers, filters are applied to the image, extracting features such as edges, lines, and shapes. Each convolutional layer typically consists of multiple filters, and each filter learns to detect a specific kind of feature in the image.

Pooling Layer (Optional): Pooling layers are frequently introduced amidst convolutional layers to diminish the data's dimensionality and regulate overfitting. Usually, pooling layers operate by condensing the results of adjacent neurons in the preceding layer. There are various pooling techniques, like average pooling and max pooling, each with slightly different effects.

Activation Layer: After each convolutional or pooling layer, there is typically an activation layer. This layer applies a non-linear transformation to the data, which helps the network learn more complex features. A common activation layer is the Rectified linear unit (ReLU) activation method.

Fully Connected Layer: After the convolutional layers, the network may have one or more fully connected layers. These layers take the output from the previous layer and connect every neuron in that layer to every neuron in the next layer. This allows the network to learn more complex relationships between the features extracted by the convolutional layers.

Output Layer: The final layer of the network is the output layer. The number of neurons in the output layer depends on the number of classes that the network is trying to classify. For example, if the network is trying to classify images of handwritten digits (0-9), then the output layer would have 10 neurons. Each neuron in the output layer corresponds to a particular class. Usually, the output layer employs a SoftMax activation function, generating a probability distribution across the classes. The highest value in this distribution indicates the class that the network is most confident the input image belongs to.

Data Gathering and Generation: Data gathering en- compasses user information collection, while data generation involves synthesizing additional profiles. Synthetic profiles are created using third-party websites and Beautiful Soup for web scraping. The process involves defining time intervals for webpage refreshes, extracting bios iteratively, and organizing them into structured DataFrames. Additional profile data, including categories like religion and politics, is generated using random numbers. Integration of these categories with the bios completes the dataset. The finalized DataFrame is exported for future use, ensuring the creation of diverse synthetic profiles. This implementation combines advanced CNN techniques for user authentication with efficient data generation strategies, enabling the creation of comprehensive synthetic profiles while
ensuring robust user authentication mechanisms.

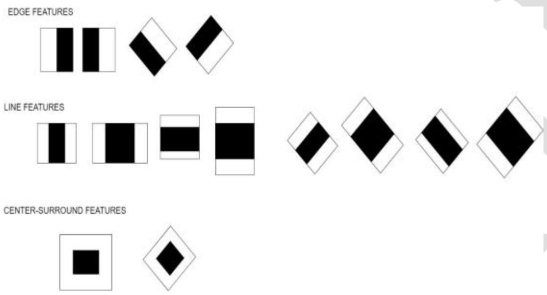


Fig. 5. Haar Cascade

Within the field of computer vision, Haar Cascade stands as a well-established machine learning algorithm for object detection. Its efficacy is rooted in a supervised learning method, wherein it undergoes training on a carefully assembled dataset comprising both affirmative and negative images. The affirmative images feature the targeted object, whereas the negative ones lack it. This thorough training regimen enables Haar Cascade develops the ability to discern subtle visual patterns characteristic of the target object.

To achieve this object recognition, Haar Cascade leverages Haar features, which function as miniature image analysts. Each Haar feature meticulously examines a specific rectangular region within a detection window. It then calculates the difference in pixel intensity between adjacent areas within that region. By systematically comparing these intensity variations across various image sections, Haar Cascade builds a robust understanding of the visual elements present.

However, the true innovation behind Haar Cascade lies in its cascading architecture. This approach can be likened to a multi-stage filtering process. The algorithm employs multiple stages, each containing increasingly complex Haar features. During detection, an image is passed through these stages sequentially. Simpler features act as a preliminary filter, efficiently discarding irrelevant regions of the image. Only promising areas containing potential object detections proceed to subsequent stages with more intricate features for a final verdict. This cascading architecture significantly enhances processing speed, making Haar Cascade particularly well-suited for real-time applications like facial recognition in video streams. While more advanced object detection techniques have been developed in recent years, Haar Cascade remains a powerful and computationally efficient tool within the computer vision community.

B. Using NLP for Text Bio

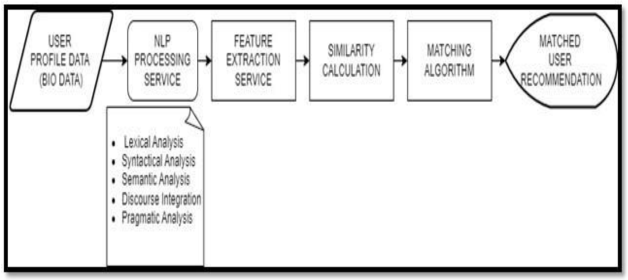


Fig. 6. NLP Flow Chart

- First, users create profiles filled with information about themselves (like bios and interests). This is called ”User profile data” or ”Bio data” in the flowchart.
- Then, a special tool called ”NLP” (Natural Language Processing) jumps in. Think of NLP as a fancy text analyzer. It goes through all that profile information and pulls out the important details, like keywords or personality traits. This process is called ”Feature extraction.”
- With those features in hand, the flowchart uses a

”Match- ing algorithm” to compare this user’s profile to everyone else’s on the platform. Like comparing ingredients in a recipe, the algorithm finds similarities between users based on the features NLP extracted. The more similar features two users have, the higher their ”Similarity score” will be.

- Finally, based on those scores, the magic happens! The flowchart uses ”Matched user recommendation” to suggest a list of potential matches for the user. These are the people with the highest similarity scores, the ones whose profiles seem most compatible according to the algorithm.

The world of matchmaking has gone digital, and with it, the power of NLP (Natural Language Processing) has emerged as a game-changer. Sure, NLP excels at the initial spark, identifying compatible profiles based on interests and personality traits gleaned from bios. But the true magic happens after that initial connection. Imagine you’ve matched with someone who, thanks to NLP analysis, shares your passion for obscure indie films. The matchmaking service, armed with this knowledge, becomes your digital cupid. It subtly guides your interaction, suggesting conversation starters that delve deeper than just ”what do you do for a living?” Think of it as a digital wingman, crafting prompts based on your unique profiles, nudging you towards discussions about your favorite directors or hidden film gems. This personalized approach, fueled by NLP, fosters deeper connections that go beyond superficial small talk.

But NLP’s role extends far beyond just sparking conversations. It acts as a communication guardian angel, analyzing your chat history to identify potential areas for further connection. For example, NLP might detect a fleeting mention of a shared love for rock climbing in your messages. Seizing this opportunity, the matchmaking service might suggest prompts or talking points related to your next climbing adventure. This subtle guidance empowers you to tailor your approach and build a stronger bond based on shared interests. It’s like having a communication coach in your pocket, constantly reminding you of the common ground you’ve discovered.

The benefits extend beyond individual interactions. By analyzing communication styles and interests extracted from user profiles, NLP can recommend alternative connections beyond the initial matches. Imagine, instead of a limited pool of options, the matchmaking service can leverage NLP to identify other users who share your communication style and express similar passions, even if their initial profiles didn’t perfectly align. This broadens the user pool and increases the chances of finding a truly compatible partner, someone who not only shares your interests but also clicks with you on a deeper conversational level.

Finally, NLP eliminates language barriers. Real-time translation powered by NLP allows you to connect with people who speak different languages. No longer will a foreign tongue stand in the way of potential love. Imagine sparking a connection with someone across the globe, all facilitated by the invisible hand of NLP.

In essence, NLP acts as the secret weapon of your digital matchmaker. It fosters meaningful communication, curates compatible connections, and ultimately increases the likelihood of forging lasting relationships. It’s a testament to the power of technology, not just in connecting people, but in helping them connect on a deeper, more meaningful level.

C. K-Mean clustering Algorithm

With K-Means, we aim to enhance matchmaking by considering these clusters, offering users personalized recommendations aligned with their distinct preferences. The implementation of this clustering technique signifies a pivotal step towards refining user experiences and optimizing the matchmaking algorithm. The application of K-Means Clustering signifies a pivotal step, representing a significant advance in comprehensively understanding user profiles. Prior to delving into cluster- ing, a meticulous data preparation process unfolds to address intricacies like missing values and feature scaling. This process culminates in determining the optimal number of clusters (K), navigated through methodologies like the Elbow Method or Silhouette Analysis. Once K is established, the K-Means algorithm takes center stage, iteratively assigning profiles to clusters based on shared features. The outcome is nuanced segmentation, with each cluster encapsulating a unique set of traits. Insights from these clusters transcend categorization, offering valuable resources for tailored recommendations, user experiences, strategic enhancements. This iterative clustering process attests to the dynamic nature of user datasets, adapting as behaviors evolve and platform dynamics shift. The evaluation phase employs metrics like the Silhouette Score to gauge clustering efficacy. This scrutiny ensures resulting clusters are distinctive and meaningful in capturing patterns within diverse profiles

- we’ve Gathered data from online datasets in the match- making platform, including attributes such as age, location, interests, hobbies, personality traits, and preferences.

- Processed the data to manage absent values, standardize numerical characteristics, and encode categorical variables if needed.
- Identified the relevant features or attributes X = x1, x2,
  ..., xn that will be used for matchmaking. These features capture the key characteristics that influence compatibility between individuals.
- K-means involves an iterative process where the data is divided into k clusters by minimizing the variance within each cluster. This process alternates between two steps:
  - Assign each data point is assigned to the cluster with the closest centroid.
  - Revise each cluster's centroid by calculating the average of all data points assigned to it.
- Mathematically, given a set of data points X = x1, x2, ..., xn and a predefined number of clusters k, K-means aims to minimize the objective function:

$$J = \sum_{i=1}^{n} \sum_{x \in C_i} ||x - \mu_i||^2$$

where Ci represents the ith cluster, µi is the centroid of cluster Ci, and ——— . ——— represents the Euclidean distance.

- Analyzed the clusters generated by K-means to under- stand the characteristics and preferences of individuals within each cluster.
- Each cluster represents a group of individuals who share similar attributes or preferences, making them potentially compatible matches for each other.
- When a user seeks matchmaking recommendations, we assigned them to the cluster that best represents their attributes by finding the nearest centroid.
- Recommends potential matches from within the same cluster to the user, as individuals within the same cluster are likely to share similar interests and preferences.
- Collect feedback from users regarding the quality of matchmaking recommendations provided.
- Use the feedback to refine the clustering algorithm and improve the accuracy of future matchmaking recommendations.
- Periodically updated the clustering model to adapt to changes in user preferences or demographic shifts.
- Re-running the K-means algorithm on updated data to generate new clusters and ensure that matchmaking recommendations remain relevant over time.

By employing K-means clustering in an AI-based matchmaking model, personalized recommendations can be made by grouping individuals with similar characteristics or preferences together, leading to more meaningful connections and im- proved user satisfaction.

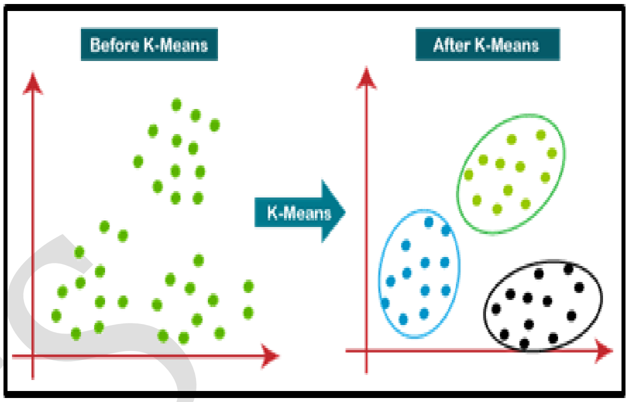


Fig. 7. K-Means Clustering

Count Vectorization and TFIDF Vectorization. To experiment and find the vectorization technique. Since our final Data Frame has over 100 features it's important to lessen its dimensionality, we employ Principal Component Analysis (PCA) to condense the dataset while retaining its features. With our data scaled, vectorized and transformed using PCA we can now proceed with the clustering process. Determining the number of clusters is crucial for grouping profiles together. We assess metrics like the Silhouette Coefficient and Davies Bouldin Score to ascertain the optimal number of clusters. These metrics offer insights into the performance of clustering algorithms. Help us make informed decisions. By running our algorithm with no. of clusters we can effectively identify, the optimal number of clusters for robust profile clustering. This approach utilizes code execution and evaluation metrics to ensure results.

D. Na¨ıve Bias recommendation system

Na¨ıve bias

We've used Na¨ıve Bias recommendation system in our model to recommend perfect matches. Naive Bayes, a probabilistic algorithm grounded in Bayes' theorem, finds extensive application in classification tasks, including recommendation systems. - The algorithm operates under the assumption that the features are unrelated to one another, thus earning the moniker ”naive.” Despite this simplification, Naive Bayes frequently demonstrates effective performance, particularly for tasks involving text classification. Naive Bayes Recommendation Bias: Despite its effectiveness, Naive Bayes recommendation systems may suffer from certain biases factors influencing the quality and fairness of recommendations may include various biases. One prevalent bias is the ”popularity bias,” where the algorithm Often suggests popular items more frequently, resulting in a shortage of diversity in recommendations. Another bias is the ”confirmation bias,” where the algorithm reinforces existing user preferences and fails to introduce users to new or diverse items that they may like. Addressing Recommendation Bias: To mitigate bias in Naive Bayes recommendation systems, various techniques can be employed: Diversification: Introducing diversity into the recommendation process by considering a wider range of item features and avoiding over-reliance on popularity metrics. Personalization: Tailoring recommendations to individual user preferences by incorporating user feedback and adapting the recommendation model over time. Fairness-aware Modeling: Incorporating fairness constraints into the recommendation model to ensure equitable treatment of users from diverse backgrounds and preferences. Evaluating Recommendation Systems: It’s essential to evaluate the performance of Naive Bayes recommendation systems to assess their effectiveness and identify potential biases. Common evaluation metrics include precision, recall, accuracy, and diversity, which measure the relevance, coverage, and fairness of recommendations.

Bayes’ Theorem calculates the likelihood of an event happening given the probability of another event that has already taken place.

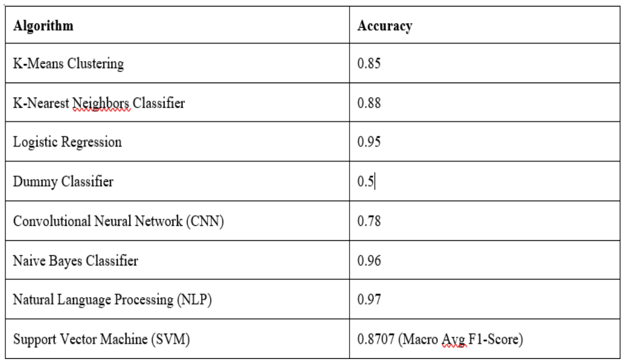
We’ve Identified the attributes or characteristics of individuals that are relevant to matchmaking. These attributes can include age, location, interests, hobbies, personality traits, preferences, etc. Gather a dataset consisting of historical data on individuals and their attributes, along with information on whether they are considered a good match or not. This dataset serves as the training data for the Naive Bayes classifier. Apply Bayes’ Theorem to calculate the probability of a match between two individuals based on their attributes. The theorem articulates that the likelihood of event A occurring given that event B has happened equals the likelihood of event B occurring given that event A has happened, multiplied by the probability of event A occurring, and then divided by the probability of event B occurring. In mathematical terms, it is expressed as:

P(A—B) = P(B—A) * P(A) / P(B)

Apply the Naive Bayes classifier, assuming that the attributes are independently conditioned on the class label (i.e., whether a match is good or not). Despite this simplifying assumption (naivety), Naive Bayes classifiers can perform well in practice, especially with large datasets. Train the Training the Naive Bayes classifier with the provided dataset. The model learns the conditional probabilities of each attribute given the class label (match or no match). Given the attributes of two individuals, use the trained Naive Bayes classifier to predict the probability of them being a good match. The classifier calculates the probable match based on the observed attributes of the individuals. Recommend matches to users based on the predicted probabilities of compatibility. Individuals with higher predicted probabilities of being a good match are recommended to each other. By employing Bayes’ Theo- rem in conjunction with the Naive Bayes classifier, AI-based matchmaking models can make personalized recommendations by analyzing the attributes and preferences of individuals, ultimately improving the matchmaking process and enhancing user satisfaction.

III. Result and Discussion

Our working model seamlessly integrates cutting edge technologies, including facial recognition, K-Means clustering, and a Naive Bayes recommendation engine. Users are securely authenticated through facial features, and their data is efficiently organized. Personalized recommendations enhance user engagement, fostering meaningful interactions. Continuous adaptation ensures responsiveness to evolving needs. In summary, our system bridges technology and human connections, creating a positive impact while prioritizing privacy, scalability, and ongoing enhancements.



Algorithm	Accuracy
K-Means Clustering	0.85
K-Nearest Neighbors Classifier	0.88
Logistic Regression	0.95
Dummy Classifier	0.5
Convolutional Neural Network (CNN)	0.78
Naive Bayes Classifier	0.96
Natural Language Processing (NLP)	0.97
Support Vector Machine (SVM)	0.8707 (Macro Avg F1-Score)

Fig. 8. Accuracy of algorithms

Future Enhancements: In the future, the system could benefit from advanced clustering algorithms, enhanced user profiling techniques, and dynamic recommendation systems adapting in real-time. Integration with social media platforms and natural language understanding for nuanced analysis of user bios are also promising avenues. Additionally, an intuitive user interface, robust privacy and security measures, and continuous optimization of machine learning models can further elevate the system's capabilities. Implementing a feedback mechanism for user input would enable refinement and improvement of recommendations over time.

Snapshots:

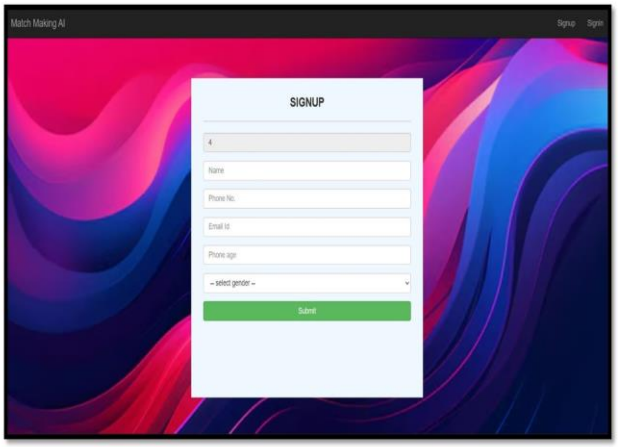
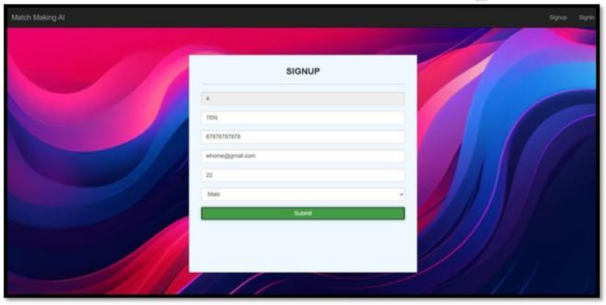


Fig. 9. Signup Page



4
TEN
67878787878
whome@gmail.com
22
Male

Fig. 10. Filled Signup page

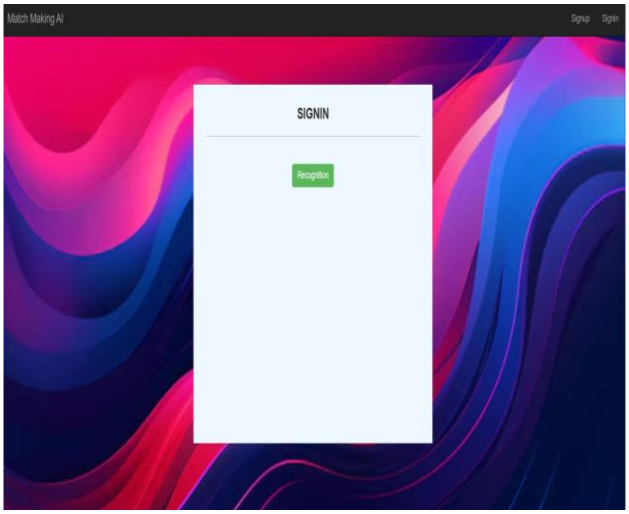


Fig. 11. SignIN Page

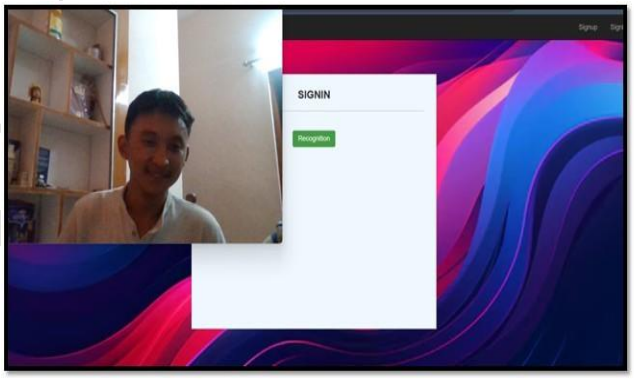


Fig. 12. Authentication

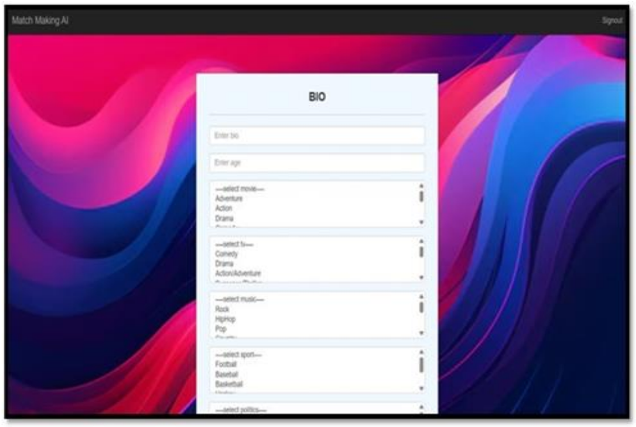
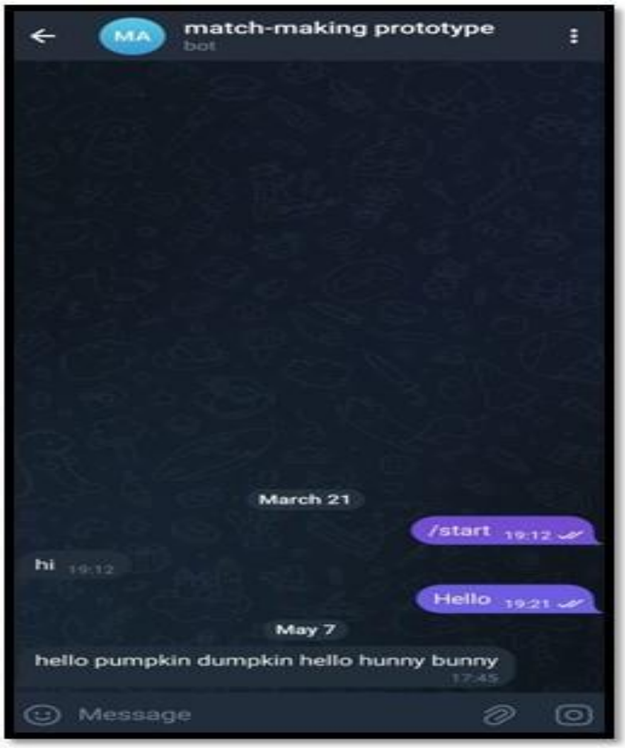


Fig. 13. Generating Bios



March 21

/start 19:12

hi 19:12

Hello 19:21

May 7

hello pumpkin dumpkin hello hunny bunny 17:45

Fig. 14. Chat Feature

IV. CONCLUSION

The integration of facial recognition for secure user authentication, K-Means clustering for efficient data organization, NLP for extracting bios, and a Naive Bayes-based recommendation system results in a robust platform aimed at enhancing user experience and fostering connections. By leveraging advanced technologies, this system ensures both security and efficiency in user interactions while providing personalized recommendations based on shared interests and preferences. Overall, it offers a seamless and enjoyable experience for users seeking meaningful connections, showcasing the potential of integrated systems in facilitating connections in online platforms.

REFERENCES

[1] Ido Guy, Naama Zwerdling, David Carmel, Inbal Ronen, Erel Uziel, Sivan Yogev, and Shila OfekKoifman. 2009. Personalized recommendation of social software items based on social relations. In Proceedings of the third ACM conference on Recommender systems (RecSys ’09). Association for Computing Machinery, New York, NY, USA, 53–60.

[2] Krzywicki, A. et al. (2010). Interaction-Based Collaborative Filtering Methods for Recommendation in Online Dating. In: Chen, L., Triantafillou, P., Suel, T. (eds) Web Information Systems Engineering – WISE 2010. WISE 2010. Lecture Notes in Computer Science, vol 6488. Springer, Berlin, Heidelberg.

[3] L. Chen, R. Nayak and Y. Xu, ”Improving Matching Process in Social Network,” 2010 IEEE International Conference on Data Mining Workshops, Sydney, NSW, Australia, 2010, pp. 305-311.

[4] L. Chen, R. Nayak and Y. Xu, ”A Recommendation Method for Online Dating Networks Based on Social Relations and Demographic Information,” 2011 International Conference on Advances in Social Networks Analysis and Mining, Kaohsiung, Taiwan, 2011, pp. 407-411.

[5] P. Xia, B. Liu, Y. Sun and C. Chen, ”Reciprocal recommendation system for online dating,” 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, France, 2015, pp. 234-241.

[6] Cao Xiao, David Mandell Freeman, and Theodore Hwa. 2015. Detecting Clusters of Fake Accounts in Online Social Networks. In Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security (AISec ’15). Association for Computing Machinery, New York, NY, USA, 91–101.

[7] R. Chen, Q. Hua, Y. -S. Chang, B. Wang, L. Zhang and X. Kong, ”A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks,” in IEEE Access, vol. 6, pp. 64301-64320, 2018.

[8] R. Fang, X. Shen, Y. Guo, J. Yao and J. Qiu, ”Switching Strategy of Recommendation Algorithms in Online Dating Platform,” 2019 Seventh International Conference on Advanced Cloud and Big Data (CBD), Suzhou, China, 2019, pp. 168-173.

[9] M. -V. Stoicescu, S. Matei and R. Rughinis, ”Sharing and Privacy in Dating Apps,” 2019 22nd International Conference on Control Systems and Computer Science (CSCS), Bucharest, Romania, 2019, pp. 432-437.

[10] Z. Cai and X. Zhang, ”Marriage Recommendation Algorithm Based on KD-KNN-LR Model,” 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 569-572.

[11] Yihan Wu and Ryan M. Kelly. 2021. Online

Dating Meets Artificial Intelligence: How the Perception of Algorithmically Generated Profile Text Impacts Attractiveness and Trust. In Proceedings of the 32nd Australian Conference on Human-Computer Interaction (OzCHI '20). Association for Computing Machinery, New York, NY, USA, 444–453.

[12] A.A. Bloshenkina and K. V. Tcyguleva, ”Application of Machine Learn- ing to Create a Recommendation in Social Communication Based on Data Analysis,” 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), St. Petersburg, Moscow, Russia, 2021, pp. 241-245.

[13] V. Vasani, T. Khanolkar, V. S. Raman, D. Kundu and T. Shah, ”Bogus User Profile Detection on Matrimonial Sites Using Machine Learning Approach,” 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2021, pp. 1-5.

[14] Eastwick, P. W., Joel, S., Carswell, K. L., Molden, D. C., Finkel,
E. J., Blozis, S. A. (2023). Predicting romantic interest during early relationship development: A preregistered investigation using machine learning. European Journal of Personality, 37(3), 276-312.

[15] N. Lutfiani, S. Wijono, U. Rahardja, A. Iriani and E. A. Nabila, ”Artificial Intelligence Based on Recommendation System for Startup Matchmaking Platform,” 2022 IEEE Creative Communication and Innovative Technology (ICCIT), Tangerang, Indonesia, 2022, pp. 1-5.

[16] Said, Yahia Barr, Mohammad Ahmed, Hossam. (2020). Design of a Face Recognition System based on Convolutional Neural Network (CNN). Engineering, Technology and Applied Science Research. 10. 5608-5612. 10.48084/etasr.3490.

[17] Duan, Jianyong Li, Ru Hu, Yi. (2009). A bio-inspired application of natural language processing: A case study in extracting multi- word expression. Expert Systems with Applications. 36. 4876-4883. 10.1016/j.eswa.2008.05.046.

[18] Saleh, Amir Dharshinni, N.Priya Perangin-Angin, Despaleri Azmi, Fadhillah Sarif, Muhammad. (2023). Implementation of Recommenda- tion Systems in Determining Learning Strategies Using the Na¨ıve Bayes Classifier Algorithm. Sinkron. 8. 256-267. 10.33395/sinkron.v8i1.11954.