Journal of International Conference Proceedings (JICP)
Vol. 8 No. 1, pp.578-591, June, 2025
E-ISSN: 2621-993X P-ISSN: 2622-0989
https://ejournal.aibpmjournals.com/index.php/JICP/index

A College Buddy System with Matching Algorithm

Chi Ling Ko[1*]
[1]School of Computing, INTI International College Penang
1-Z, Lebuh Bukit Jambul, Bukit Jambul, 11900 Bayan Lepas, Pulau Pinang
*Corresponding Email: p21013435@student.newinti.edu.my

ARTICLE INFORMATION

Publication information

Research article

HOW TO CITE
Ko, C. L. (2025). A college buddy system with matching algorithm. *Journal of International Conference Proceedings*, *8*(1), 578–591.

DOI:
https://doi.org/10.32535/jicp.v8i1.4001

Received: 26 April 2025
Accepted: 27 May 2025
Published: 28 June 2025

ABSTRACT

This study introduces *BuddyIN*, a mobile application developed to strengthen peer mentorship and collaborative learning among students at INTI International College Penang. The primary objective is to provide an intelligent platform that efficiently matches students with suitable study partners while offering academic support features. The system was developed using the Waterfall methodology, encompassing requirement gathering, system design, implementation, testing, deployment, and maintenance. Key components include the buddy matching algorithm, AI chatbot integration, intuitive user interface design, and extensive system testing to ensure functionality and reliability. The matching algorithm effectively pairs senior students with freshmen based on user-input data, fostering meaningful academic connections. The project successfully produced a fully functional mobile application that promotes productive study partnerships, delivers real-time academic assistance, and enhances students' learning experiences, time management, and overall academic performance.

Keywords: Academic Assistance; Artificial Intelligence; Buddy Matching Algorithm; Buddy System; Collaborative Learning; Mobile Application

INTRODUCTION

The Buddy System is a program designed to match freshmen with older and more experienced students who serve as mentors, offering guidance and practical advice to help new students adapt to university life (Eduopinions, 2022). Prior research has shown that the buddy method for peer mentorship is highly effective in enhancing knowledge and skills, particularly in the context of scientific research (Balan, 2015). By engaging in this system, students are able to exchange ideas from multiple perspectives, thus enriching their academic and social experiences.

At INTI International College Penang, significant challenges remain in providing effective peer mentorship and collaborative learning opportunities. Currently, there is no structured platform that efficiently connects freshmen with senior peers for academic guidance, practical advice, and campus orientation. This gap particularly affects international students, who often find it difficult to build meaningful friendships with local peers. As highlighted in earlier studies, buddy programs help international students better understand host country culture and foster a sense of belonging (Nilsson, 2019). Moreover, statistical evidence indicates that peer assistance is a vital source of academic support. According to Holland & Pawlikowska, (2019), 52 out of 141 students reported that their second choice in seeking help with academic difficulties was to approach a peer or study buddy. This finding underscores the crucial role that peer mentors play in supporting students' learning and adjustment to campus life. Without a dedicated platform to facilitate these mentorship connections, however, freshmen may face greater challenges in navigating university life and achieving their full academic potential (Naidoo et al., 2021).

To address these challenges, this study proposes the development of a mobile application to support students at INTI International College Penang in finding study partners, forming study groups, and engaging in peer-to-peer collaboration. The application incorporates features such as matching students based on shared interests or courses, thus enabling personalized study processes through AI-based matching (Oh et al., 2025). Peer support has been found to positively influence performance and self-confidence, especially when facilitated by experienced students (Wahyuni, 2022). Additionally, the app provides a platform for group discussions, integrates an AI chatbot capable of answering academic questions, and allows both students and administrators to upload and share study resources. This comprehensive approach aims to foster collaborative learning, strengthen peer mentorship, and enhance academic outcomes for students.

LITERATURE REVIEW

Evolution of the Buddy System

The buddy system has evolved significantly in response to shifts in educational contexts and technological advancements. Prior to 2020, the system primarily relied on face-to-face interactions, where senior students mentored freshmen to support their academic and social integration into university life. During the COVID-19 pandemic (2020–2021), educational institutions transitioned to online platforms such as Zoom and Microsoft Teams to maintain mentorship and peer support virtually (Andries & Lengkoan, 2023). In the post-pandemic period (2022–May 2024), the system adopted a hybrid model that combined physical and virtual platforms, thereby increasing inclusivity and adaptability. Looking forward, the integration of advanced technologies—including artificial intelligence (AI) and matching algorithms—is expected to further enhance the accuracy and effectiveness of buddy pairings by incorporating factors such as course compatibility, personal interests, and personality traits.

Matching Algorithms

Matching algorithms are essential tools for pairing individuals or items based on similarity or compatibility. They play a vital role in applications such as recommendation systems, peer-to-peer networks, and buddy matching systems. These algorithms operate by calculating similarity values between subjects and assigning matches according to established criteria of proximity (Rosenbaum, 2001).

K-Nearest Neighbours (KNN)

The K-Nearest Neighbours (KNN) algorithm is a supervised machine learning method applied in both regression and classification problems. As an instance-based learning technique, KNN classifies a data point according to the majority class of its k nearest neighbours. It is simple, flexible, and does not require assumptions about the underlying data distribution. However, it can be computationally expensive for large datasets and sensitive to irrelevant features and noise (Zhang et al., 2024). KNN works by first calculating the distance between a new data point and every point in the training dataset, then assigning the new point to the majority class among the closest neighbours (GeeksforGeeks, 2024a).

Decision Trees

Decision Trees are tree-structured classifiers where nodes represent features, edges represent decisions, and leaves represent outcomes. They are widely used for both classification and regression tasks due to their interpretability and intuitive visualization. Decision Trees can handle categorical and numerical data but are prone to overfitting, particularly with large or complex datasets. They may also become biased if certain classes dominate the training data. To improve their robustness and generalization, techniques such as feature selection, pruning, and ensemble approaches (e.g., Random Forests) are often applied (Rahman et al., 2024). Decision Trees function by recursively selecting the most informative feature to partition the dataset until a stopping criterion is met, producing predicted values or class labels at the leaves (Abhis, 2024).

Support Vector Machines (SVM)

Support Vector Machines are supervised learning models that determine the optimal hyperplane in a high-dimensional space to separate data classes. SVM belongs to the generalized linear classification family and is applicable for both regression and classification. It performs effectively in high-dimensional spaces and when there is a clear margin of separation between classes (Jair et al., 2020). Based on the principle of Structural Risk Minimization (SRM), SVM maximizes the margin between support vectors while minimizing classification errors. Kernel functions such as linear, polynomial, and radial basis function (RBF) are commonly employed to map input data into higher-dimensional spaces, enabling the classification of non-linearly separable data (Aswathisasidharan, 2023).

Naïve Bayes

Naïve Bayes is a probabilistic classifier grounded in Bayes' theorem, which assumes independence among features. Despite this simplifying assumption, it has been widely applied due to its computational efficiency and effectiveness in tasks such as text categorization and spam filtering (Li & Li, 2020). For small to medium-sized datasets, Naïve Bayes often performs remarkably well. Its formula is given as:

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (1)$$

where, P(H|E) = Probability's positional probability (conditional probability)
P(E|H) = Parameter E's probability in relation to hypothesis H
P(H) = The prior hypothesis, or prior probability (Peling et al., 2017)

The process begins with training data to calculate prior probabilities for each class. Probabilities for each feature within each class are then estimated using probability density functions or histograms. Finally, the theorem is applied to predict the most likely class for new data, assigning it to the class with the highest posterior probability (GeeksforGeeks, 2024b).

Distance Calculations for the KNN Model
Several distance measures are commonly used in the K-Nearest Neighbours (KNN) model, each with its own strengths and limitations depending on the type of dataset.

Cosine Similarity
Cosine similarity measures the cosine of the angle between two non-zero vectors in an inner product space, making it particularly useful for complex data such as text. It evaluates whether two vectors are pointing in the same direction rather than focusing on their magnitude (Han et al.,2012). Because it emphasizes direction over magnitude, cosine similarity is effective in text analysis and document similarity tasks (Marakani, 2019). The formula is:

$$Cosine\ Similarity = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad (2)$$

where A and B are the vectors being compared, A_i and B_i are the components of these vectors (Varun, 2020).

Euclidean Distance
Euclidean distance calculates the straight-line distance between two points in Euclidean space, derived from the Pythagorean theorem. It is widely applied in machine learning, particularly for continuous variables and simple datasets, due to its simplicity and intuitiveness (Marakani, 2019; Suwanda et al., 2020). However, it is sensitive to data volume and scaling issues, which may reduce its effectiveness in high-dimensional datasets. The formula is:

$$d(A,B) = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \quad (3)$$

where A_i and B_i are the coordinates of the points in each dimension (Suwanda et al., 2020)

Gower Distance
Gower distance is a dissimilarity measure capable of handling mixed data types, including numerical, categorical, and binary variables. This flexibility makes it suitable for heterogeneous datasets. However, it requires careful weighting of variables to ensure

Journal of International Conference Proceedings (JICP)
Vol. 8 No. 1, pp.578-591, June, 2025
E-ISSN: 2621-993X P-ISSN: 2622-0989
https://ejournal.aibpmjournals.com/index.php/JICP/index

meaningful representation, and it can be computationally intensive (D'Orazio, 2021; Vieira, 2023). The formula is:

$$S_{\text{Gower}}(x_i, x_j) = \frac{\sum_{k=1}^{p} s_{ijk}\delta_{ijk}}{\sum_{k=1}^{p} \delta_{ijk}}. \tag{4}$$

where δ_{ijt} is 1 if the comparison is valid, 0 otherwise. There will be different calculations method for numerical variables, binary variables and categorical variables (Vieira, 2023).

Jaccard Distance

Jaccard distance measures dissimilarity between two sets by comparing the size of their intersection relative to their union. It is frequently applied in text similarity, clustering, social network analysis, and recommender systems (Kabasakal & Soyuer, 2021). Jaccard distance works particularly well with binary and categorical data, though it is less effective for continuous data or sets with many overlapping elements. The formula is:

$$d(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \tag{5}$$

where $|A \cap B|$ is the size of the intersection of the sets and $|A \cup B|$ is the size of the union of the sets (Lehal, 2017).

Selection Considerations for KNN

The choice of distance metric plays a crucial role in determining the effectiveness of the K-Nearest Neighbours (KNN) model. Although KNN is widely appreciated for its simplicity, flexibility, and ability to handle both classification and regression without an explicit training phase, its performance is highly dependent on the suitability of the distance measure applied.

Cosine similarity is generally preferred when dealing with text or vector-based data, as it emphasizes the orientation of vectors rather than their magnitude, making it useful in contexts where direction is more meaningful than scale. In contrast, Euclidean distance remains the most commonly used metric for continuous and low-dimensional datasets because of its straightforward computation and intuitive interpretation, although it tends to be less reliable in high-dimensional spaces. Gower distance offers significant advantages in handling heterogeneous datasets containing numerical, categorical, and binary variables, providing adaptability in complex data scenarios. However, it demands higher computational resources and careful adjustment of variable weights to ensure balanced representation. Jaccard distance, on the other hand, is particularly well-suited for binary and categorical data, especially when the focus is on measuring similarity or dissimilarity between sets.

Among these options, Gower distance is often selected for datasets that combine multiple variable types, as it enhances both the robustness and accuracy of KNN models. Despite its computational complexity and the need for normalization, it is considered a versatile and reliable metric when addressing the challenges of heterogeneous data.

Comparison between the Existing System and the Proposed System

Table 1. Differences Between The Existing System and The Proposed System

Features	Navigate Student	Course Hero	Studocu	Quora	Proposed System
Study Buddies / Buddy Matching	Yes, but lack of personalise buddy matching	No	No	No	Personalise the buddy matching
Document Sharing	No	Yes	Yes	No	Yes
Discussion Forum	No	No	No	Yes	Yes
Community Interaction	Limited to student interactions	Limited to document sharing and Q&A	Limited to document sharing and Q&A	Q&A, comments, upvotes/ downvotes	Document sharing, Q&A and comments
AI chatbots	No	Yes	Yes	Yes	Yes

Note. Adapted from (Quora.Inc, 2012), (Global, 2014), (StuDocu, 2019), (Hero, 2015)

The proposed system addresses the limitations of existing platforms, as highlighted in Table 1, by introducing a personalised buddy matching feature, unrestricted document sharing, and a discussion forum similar to Quora that encourages collaborative learning. In addition, the system incorporates an AI-powered chatbot tailored specifically for INTI students, providing instant academic assistance and creating a more comprehensive and supportive learning environment. By integrating these features, the proposed system bridges the gaps left by current solutions and delivers a more holistic approach to peer learning and academic support.

RESEARCH METHOD

This project employed the Waterfall methodology, a structured software development life cycle (SDLC) model characterized by sequential phases. The process began with requirement gathering, which was conducted through a comprehensive literature review, analysis of existing systems, and evaluation of relevant technologies. This stage provided a clear foundation for the subsequent design phase.

During the system design phase, the architecture of the proposed application was outlined using UML diagrams (see Appendix A), database schemas, and user interface mock-ups created with tools such as Lucidchart and Figma. The implementation stage involved developing the mobile application using Android Studio, with backend integration of key functionalities including the K-Nearest Neighbour (KNN) algorithm for personalised buddy matching and the Gemini model for chatbot support.

The testing phase was conducted to ensure that both functional and non-functional requirements were met. This was achieved through unit testing, integration testing, and performance testing, which verified the reliability and efficiency of the system. Finally, the deployment and maintenance phase involved demonstrating the system on Android devices and publishing the project on GitHub. Provisions for ongoing maintenance, such as updates and issue resolution, were also incorporated to ensure the long-term stability and usability of the application.

RESULTS

The buddy matching system was implemented using a Flask server integrated with Firebase Realtime Database and developed based on the K-Nearest Neighbours (KNN) algorithm with Gower distance for similarity computation. The algorithm effectively fetched and pre-processed user data including course, seniority, hobbies, and personalities. After label encoding and formatting the input data, similarity matrices were computed to identify the top 3 most similar buddies for each user.

Figure 1. Overview Result for Combined Buddies

```
Similarity Matrix:
uid                          8I1zec5JckevXHzKxhNmD6tZGCz2  Bs6bLmgT1NOzcMkUns6JNwGoUa72  ...  r0ADVXTrbZQ8laakehsv9O3WsNz1  X0Il44rOcVdTkFJlBFgDMZMoQQR2
uid                                                                                    ...
8I1zec5JckevXHzKxhNmD6tZGCz2                      0.000000                      0.321429  ...                      0.142857                      0.482143
Bs6bLmgT1NOzcMkUns6JNwGoUa72                      0.321429                      0.000000  ...                      0.250000                      0.232143
H3mTUxYVrEZGCmcFtBjJXTH2P1z1                      0.267857                      0.446429  ...                      0.196429                      0.607143
IOzfEu3xRWYDA8rPIyNTRvYqZUy2                      0.410714                      0.375000  ...                      0.267857                      0.535714
SlQxfLITEcOzGKPxSq9Y7APXPQo1                      0.375000                      0.267857  ...                      0.303571                      0.357143
qv2XqgAj1ag75I9N76WA21JSydu2                      0.071429                      0.250000  ...                      0.142857                      0.410714
r0ADVXTrbZQ8laakehsv9O3WsNz1                      0.142857                      0.250000  ...                      0.000000                      0.410714
X0Il44rOcVdTkFJlBFgDMZMoQQR2                      0.482143                      0.232143  ...                      0.410714                      0.000000

[8 rows x 8 columns]
Distances: [[0.         0.07142857 0.14285715]
 [0.         0.23214287 0.25      ]
 [0.         0.14285715 0.19642858]
 [0.         0.14285715 0.25      ]
 [0.         0.25       0.25      ]
 [0.         0.07142857 0.14285715]
 [0.         0.14285715 0.14285715]
 [0.         0.23214287 0.35714287]]
Indices: [[0 5 6]
 [1 7 5]
 [2 3 6]
 [3 2 4]
 [4 3 2]
 [5 0 6]
 [6 0 5]
 [7 1 4]]
{'X8FHufhudTNt9TXB8jy7OjJpqQH3': ['Bs6bLmgT1NOzcMkUns6JNwGoUa72', 'SlQxfLITEcOzGKPxSq9Y7APXPQo1']}
```

Figure 2. Overview Result for Other Buddies

```
Similarity Matrix:
uid                          FnVSFNdxz9g6bjDVNTF1Itor0Li2  KH08afZZooQ2Yrf8ngZT4o6qlrC3  ...  yQmaWhQtYfdJH2PzmSzTIGCK8yb2  X8FHufhudTNt9TXB8jy7OjJpqQH3
uid                                                                                    ...
FnVSFNdxz9g6bjDVNTF1Itor0Li2                      0.000000                      0.157197  ...                      0.486742                      0.413258
KH08afZZooQ2Yrf8ngZT4o6qlrC3                      0.157197                      0.000000  ...                      0.462121                      0.297727
OLLSHhqIQeZOStFFEXsmMMlOraI3                      0.291667                      0.176136  ...                      0.554924                      0.303409
Uo5Q4NGnqtYaHt3Cuqj2ZON78Cs1                      0.389015                      0.409848  ...                      0.347727                      0.357576
VaPHVJZiWJTDKfKpKTznURAN2973                      0.664773                      0.549242  ...                      0.178030                      0.334848
WcjD4lpBgqM9YKw88ItrJyDVIqi1                      0.256061                      0.326136  ...                      0.397348                      0.323864
X0Il44rOcVdTkFJlBFgDMZMoQQR2                      0.479545                      0.364015  ...                      0.442803                      0.066288
c52zsXvEq5OA0PubrqiFuKP8lw62                      0.478788                      0.545076  ...                      0.382955                      0.434470
jpdef4kROfV7zpQsfDuhgImBqZu1                      0.501136                      0.389394  ...                      0.156061                      0.387121
s70tByAwk2TU9cMRc6jvafJS5uG2                      0.660985                      0.503788  ...                      0.215909                      0.456061
tIfG191P4UStS7aOElVUL5CfBpi1                      0.255303                      0.143561  ...                      0.360227                      0.241288
yQmaWhQtYfdJH2PzmSzTIGCK8yb2                      0.486742                      0.462121  ...                      0.000000                      0.376515
X8FHufhudTNt9TXB8jy7OjJpqQH3                      0.413258                      0.297727  ...                      0.376515                      0.000000

[13 rows x 13 columns]
Distances: [[0.         0.15719697 0.25530303]
 [0.         0.1435606  0.15719697]
 [0.         0.17613636 0.27878788]
 [0.         0.2602273  0.34772727]
 [0.         0.1780303  0.21212122]
 [0.         0.18257576 0.2560606 ]
 [0.         0.06628788 0.27878788]
 [0.         0.2602273  0.38295457]
 [0.         0.1560606  0.15984848]
 [0.         0.15984848 0.21212122]
 [0.         0.1435606  0.18257576]
 [0.         0.1560606  0.1780303 ]
 [0.         0.06628788 0.24128787]]
```

Journal of International Conference Proceedings (JICP)
Vol. 8 No. 1, pp.578-591, June, 2025
E-ISSN: 2621-993X P-ISSN: 2622-0989
https://ejournal.aibpmjournals.com/index.php/JICP/index

```
Indices: [[ 0  1 10]
 [ 1 10  0]
 [ 2  1  6]
 [ 3  7 11]
 [ 4 11  9]
 [ 5 10  0]
 [ 6 12  2]
 [ 7  3 11]
 [ 8 11  9]
 [ 9  8  4]
 [10  1  5]
 [11  8  4]
 [12  6 10]]
{'X8FHufhudTNt9TXB8jy7OjJpqQH3': ['X0Il44rOcVdTkFJlBFgDMZMoQQR2', 'tIfG191P4UStS7aOElVUL5CfBpi1']}
192.168.0.120 - - [23/Oct/2024 23:37:44] "POST /get_buddies HTTP/1.1" 200 -
```

Figures 1 and Figure 2 illustrate the matching results categorised into two groups. Figure 1 displays the combined buddy list, which matches the users with the same course but different seniority levels. Figure 2 shows that the other buddy list, which matches the users, is based on general attributes such as hobbies, personalities, and seniority when the course did not match. The results are displayed in a decimal similarity score. The similarity score has smaller values means the buddy and user have similarity between each other, which is a Gower Distance characteristic that calculate the dissimilarity between objects. The matching result will be shown in the last lines with the matched buddy's id for the combined buddies who are same course with higher or lower seniority. The results will be stored under Firebase and checked for duplications before storing data. The system successfully demonstrated its ability to return personalised and course-relevant buddy suggestions in real-time using the trained model.

Figure 3. User Registration

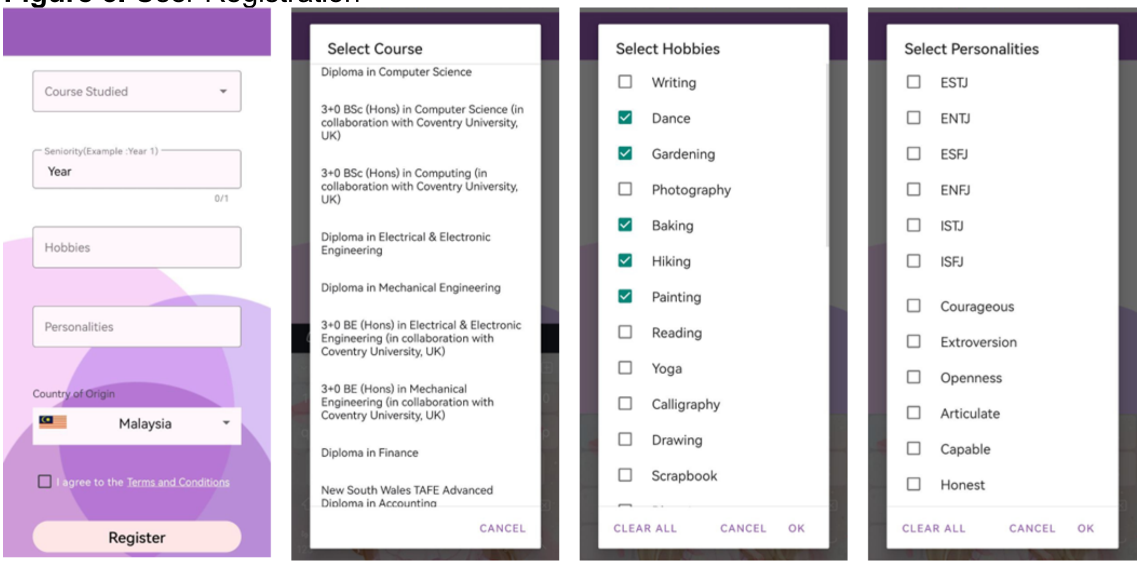
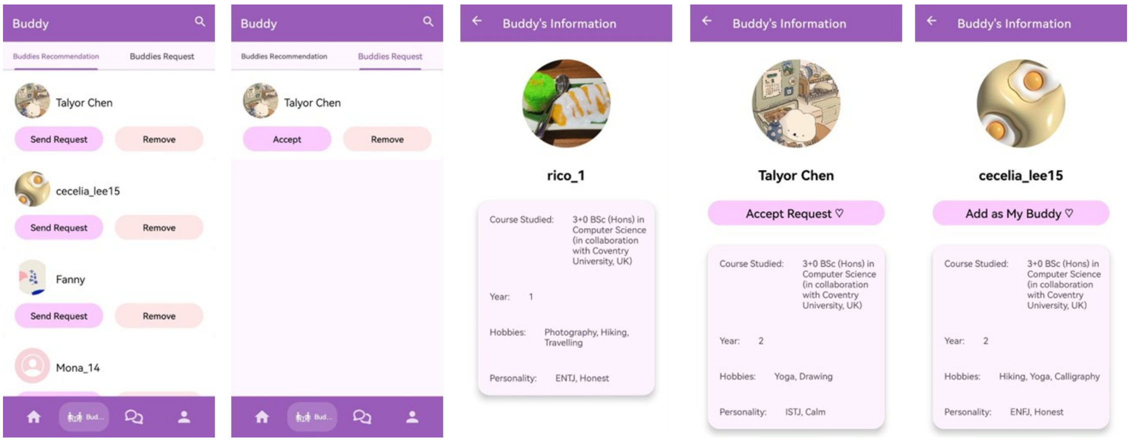


Figure 3 above shows the sign-up page for the student role and the personal information filling page for the matching algorithm's purpose. The user needs to fill up all the information field to complete the account creation for the application. After the account creation, the user will be able to match the buddies in the app.

Figure 4. Buddy Recommendation

Journal of International Conference Proceedings (JICP)
Vol. 8 No. 1, pp.578-591, June, 2025
E-ISSN: 2621-993X P-ISSN: 2622-0989
https://ejournal.aibpmjournals.com/index.php/JICP/index



The Figure 4 above illustrate the buddy recommendation page of the BuddyIN application. Users can send or remove requests to matched buddies and view detailed profiles. User can also accept or decline incoming buddy requests. The profile page displays key information used in the matching algorithm, such as course, academic year, personality, and hobbies. Additionally, users have the option to add buddies directly from the profile page.

An accuracy of the algorithm evaluation was performed using dataset of user data to evaluate the effectiveness of the buddy matching algorithm. The dataset included the essential attributes such as course, seniority level, hobbies, and personalities. These attributes were pre-processed and encoded for compatibility with the K-Nearest Neighbours (KNN) model using the Gower distance metric, which is particularly suited for mixed data types. The model was evaluated using 5-Fold Cross-Validation, which splits the dataset into five equal parts, training the model on four parts and validating it on the remaining part in each iteration. This technique provides a reliable estimate of the model's ability by ensuring every data point is used for testing.

Table 2. Average Results for the 5-Fold Cross-Validation

Metrics	Value
Accuracy	0.78
Precision	0.80
Recall	0.93
F1 score	0.84

Table 2 above shows the results of the accuracy testing of the K-Nearest Neighbours (KNN) classifier with Gower distance using cross-validation, focusing on accuracy, precision, recall, and F1 score. The accuracy score is 0.78, indicating 78% of all buddy match predictions made by the model were correct. This metric is useful for the overall but would be misleading if the datasets were imbalanced. Precision score measures the number of true matches among all match predictions. Precision score is 0.80, indicating 80% of positive predictions are correct. A high precision score is especially important when false matches are costly, such as wasting user time or degrading user experience. Recall score is 0.93, indicating 93% of positive instances are correctly identified. The high recall means the model captured almost every real buddy match. This is particularly beneficial in social matching contexts, as stopping to suggest a suitable buddy (false negative) is generally worse than incorrectly suggesting one (false positive). F1 score is the harmonic mean of precision and recall, indicating a good balance between precision and recall, which having 84%. A high F1 score reflects the model's strong ability to balance sensitivity (recall) and specificity (precision). This has proven that the model is well-developed for recommendation accuracy in buddy matching. The combined score (accuracy, precision, recall, F1) is 83.95%, suggesting the model performs well across all metrics.

Table 3. Confusion Matrix

	Predicted: Match	Predicted: Not Match
Actual: Match	14 (TP)	0(TN)
Actual: Not Match	1(FP)	7(FN)

Note. TP = True Positive, FN = False Positive, TN = True Negative, FN = False Negative

Confusion Matrix was carried out using the prediction across the fold for further understanding of the classification performance. The confusion matrix shows the 14 true positives, 1 false positive, 7 true negatives, and 0 false negative. The model correctly predicted the suitable buddy pairs with 14 true positive and 0 false negative buddy pairs, which failed to pair an actual buddy, which shows the model is highly sensitive and precise. The results show the algorithms never missed a true buddy match, which is an important condition for a buddy recommendation system to ensure no potential compatible pair is missed. Overall, the K-NN classifier's performance is evaluated using cross-validation, providing a comprehensive evaluation of its effectiveness.

DISCUSSION

This study shows that a K nearest neighbours (KNN) matching approach using Gower distance can generate useful peer matching recommendations in a heterogeneous real world student dataset. Using fivefold cross validation, the model achieved Accuracy 0.78, Precision 0.80, Recall 0.93, and F1 0.84, indicating a good balance between correctly identifying compatible buddies and limiting incorrect suggestions. The confusion matrix indicates no false negatives and only one false positive, which is desirable in a social matching context because missing a truly compatible peer is costlier than suggesting an extra candidate. These results suggest that algorithmic matching can meaningfully support peer mentoring and study partnerships in higher education.

Methodologically, Gower distance is appropriate for BuddyIN because it handles mixed attributes such as courses (categorical), seniority (ordinal), hobbies, and personality traits (categorical or binary) within a single similarity framework (D'Orazio, 2021; Vieira, 2023). By contrast, Euclidean distance assumes continuous and similarly scaled features and tends to degrade with high dimensionality (Suwanda et al., 2020). Cosine similarity focuses on directional alignment of vectors, which is effective for text but less natural for mixed type profiles (Han et al., 2012; Varun, 2020). Jaccard distance works well for set based or binary data but discards magnitude information that can matter for multiattribute matching (Kabasakal & Soyuer, 2021; Lehal, 2017). Within the KNN family, our choice reflects a fit for purpose metric selection rather than algorithmic complexity for its own sake (GeeksforGeeks, 2024a).

Substantively, the findings align with prior literature showing that buddy and peer systems enhance knowledge acquisition, adjustment, and academic confidence, especially for freshmen and international students (Balan, 2015; Nilsson, 2019; Naidoo et al., 2021). Survey evidence also indicates that students frequently seek help from peers or study buddies for academic difficulties (Holland & Pawlikowska, 2019). BuddyIN operationalises these insights by providing course aware and interest aware matching together with a discussion forum and an AI chatbot that enables just in time academic support (Oh et al., 2025).

Compared with existing platforms such as Navigate Student (Global E., 2014), Course Hero (Hero, 2015), StuDocu (2019), and Quora (Quora Inc., 2012), BuddyIN's contribution is the personalised buddy matching core. Competing services emphasise student services, document repositories, or question and answer communities, but they do not natively perform multiattribute and course sensitive pairing for mentorship and study partnership. BuddyIN complements these ecosystems by combining targeted matching, resource sharing, community discussion, and AI assistance in one application.

Several enhancements are warranted. Larger and more diverse datasets should be collected to assess generalisability beyond a single campus population, and future work should include qualitative user feedback on match satisfaction and perceived relevance to triangulate the quantitative metrics. Deploying the backend to a reliable cloud environment would improve uptime and support scale. Exploring hybrid recommenders, for example KNN with explainable rules or model stacking, and adding explanations for why a buddy was recommended may increase user trust and adoption (Rahman et al., 2024; Zhang et al., 2024). Finally, offering rematch options and stronger account and data controls will support privacy, autonomy, and long-term engagement.

CONCLUSION

In conclusion, BuddyIN is a mobile application that can foster the productivity of studies by having a partnership, and real-time academic support with lecturer assistance to enhance students' learning experience and academic success. The objectives of the project have been successfully achieved throughout the development process, such as matching algorithms with K-Nearest Neighbours (KNN) and Gower Distance, implementing a discussion platform for knowledge sharing, and including an AI-powered chatbot for support. From overcoming technical challenges to implementing machine learning models and APIs, this project also offered a wealth of learning opportunities. Every obstacle necessitated a combination of technical flexibility, analytical reasoning, and tenacity, which eventually aided in the project's accomplishment and the individual's development. The BuddyIN application can bring up benefits to students and lecturers at INTI International College Penang by supporting student success with the features in the system.

LIMITATION

This study is subject to several limitations. Firstly, the dataset was moderate and limited in scope, comprising only data related to students from INTI International College Penang. This restriction may reduce the diversity of user attributes and limit the model's ability to generalize to broader or more diverse populations. Secondly, the matching algorithm's effectiveness in capturing more complex social or behavioural dynamics has not been fully explored. Additionally, the deployment of the Flask server remains constrained, as it is hosted locally and cannot maintain continuous cloud-based activity, which limits the system's scalability and availability for real-time usage.

ACKNOWLEDGMENT

The author expresses sincere gratitude to all who supported this project. Appreciation is extended to family for their encouragement, to lecturers and the supervisor for their invaluable guidance, and to friends for their technical support and constructive feedback.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declare no conflicts of interest related to the research, authorship, or publication of this project. The study was conducted independently, free from commercial or financial influence, with all findings based solely on academic inquiry and interest in improving peer support systems. Any future conflicts will be disclosed to maintain transparency and integrity.

REFERENCES

Abhis. (2024, May 30). *Python | Decision tree implementation*. GeeksforGeeks. https://www.geeksforgeeks.org/decision-tree-implementation-python/

Andries, F., & Lengkoan, F. (2023). The importance of students' perception of online learning during pandemic. *International Journal of Applied Business and International Management, 8*(2), 294–305. https://doi.org/10.32535/ijabim.v8i2.2475

Aswathisasidharan. (2023, May 30). *Support vector machine (SVM) algorithm*. GeeksforGeeks. https://www.geeksforgeeks.org/support-vector-machine-algorithm/

Balan, H. R. (2015). Using peer mentoring-buddy system as an intervention strategy to enhance science research skills. *Journal of Multidisciplinary Studies, 4*(1), 168-185. http://dx.doi.org/10.7828/jmds.v4i1.850

D'Orazio, M. (2021). Distances with mixed type variables, some modified Gower's coefficients. arXiv. *arXiv preprint arXiv:2101.02481*.

Eduopinions. (2022, June 2). *Why is a buddy system so important at university?* Eduopinions. https://www.eduopinions.com/blog/student-life/why-is-a-buddy-system-so-important-at-university/

GeeksforGeeks. (2024a, May 30). *K-nearest neighbor (KNN) algorithm*. GeeksforGeeks. https://www.geeksforgeeks.org/k-nearest-neighbours/

GeeksforGeeks. (2024b, May 30). *Naive Bayes classifiers*. GeeksforGeeks. https://www.geeksforgeeks.org/naive-bayes-classifiers/

Global E. (2014). *Navigate student* [Mobile app]. Google Play. https://play.google.com/store/apps/details?id=com.eab.se&hl=en&gl=US

Han, J., Kamber, M., & Pei, J. (2012). Getting to know your data. In *Data mining* (pp. 39-82). Elsevier. https://doi.org/10.1016/B978-0-12-381479-1.00002-2

Hero, C. (2015). *Course Hero: AI homework help* [Mobile app]. Google Play. https://play.google.com/store/apps/details?id=com.coursehero.coursehero&hl=en&gl=US

Holland, J. C., & Pawlikowska, T. (2019). Learning Cardiac Embryology—Which Resources Do Students Use, and Why?. *Medical Science Educator*, *29*(4), 1051-1060. https://doi.org/10.1007/s40670-019-00803-4

Jair, C., García-López, E., Méndez, M., & Fernández, G. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing, 408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Kabasakal, İ., & Soyuer, H. (2021, March). A Jaccard similarity-based model to match stakeholders for collaboration in an industry-driven portal. In *Proceedings* (Vol. 74, No. 1, p. 15). MDPI. https://doi.org/10.3390/proceedings2021074015

Lehal, M. S. (2017). Comparison of cosine, Euclidean distance and Jaccard distance. *Int J Sci Res Sci, Eng Technol (IJSRSET)*, *3*(8), 1376-1381.

Li, Q. N., & Li, T. H. (2020). Research on the application of Naive Bayes and Support Vector Machine algorithm on exercises Classification. In *Journal of Physics: Conference Series* (Vol. 1437, No. 1, p. 012071). IOP Publishing. https://doi.org/10.1088/1742-6596/1437/1/012071

Marakani, S. (2019). Employee matching using machine learning. (Master's thesis, Blekinge Institute of Technology).

Naidoo, A., Byles, H., & Kwenaite, S. (2021). Student support and transition through a buddy programme to foster social integration. *Journal of Student Affairs in Africa*, *9*(2), 47-63.

Nilsson, P. A. (2019). The buddy programme: Integration and social support for international students. *Journal of Comparative & International Higher Education, 11*(Winter), 38–43. *https://doi.org/10.32674/jcihe.v11iWinter.1095*

Oh, Z. J., Odebunmi, A. T., Tu, G. Y., Teoh, Y. E., Thoe, W. M., Toh, Y. J., & Kee, D. M. H. (2025). The influence of artificial intelligence on customer retention in the Asian e-commerce market. *Asia Pacific Journal of Management and Education, 8*(1), 170–186. https://doi.org/10.32535/apjme.v8i1.3865

Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of data mining to predict period of students study using Naive Bayes algorithm. *International Journal of Engineering and Emerging Technology, 2*(1), 53-57. https://doi.org/10.24843/IJEET.2017.v02.i01.p11

Quora Inc. (2012). *Quora: The knowledge platform* [Mobile app]. Google Play. https://play.google.com/store/apps/details?id=com.quora.android&hl=en&gl=US

Rahman, M., Kamal, N., & Abdullah, N. F. (2024). EDT-STACK: A stacking ensemble-based decision trees algorithm for tire tread depth condition classification. *Results in Engineering*, *22*, 102218. https://doi.org/10.1016/j.rineng.2024.102218

Rosenbaum, P. (2001). Observational studies: Overview. In *International encyclopedia of the social & behavioral sciences* (pp. 10808–10815). Elsevier. https://doi.org/10.1016/B0-08-043076-7/00483-6

StuDocu. (2019). *Studocu: Study notes & sharing* [Mobile app]. Google Play. https://play.google.com/store/apps/details?id=com.studocu.app&hl=en&gl=US

Suwanda, R., Syahputra, Z., & Zamzami, E. M. (2020, June). Analysis of euclidean distance and manhattan distance in the K-means algorithm for variations number of centroid K. In *Journal of Physics: Conference Series* (Vol. 1566, No. 1, p. 012058). IOP Publishing. https://doi.org/10.1088/1742-6596/1566/1/012058

Varun. (2020, September 27). *Cosine similarity: How does it measure the similarity, maths behind and usage in Python*. Towards Data Science. https://towardsdatascience.com/cosine-similarity-how-does-it-measure-the-similarity-maths-behind-and-usage-in-python-50ad30aad7db

Vieira, R. (2023, May 20). *Gower distance*. ruivieira.dev. https://ruivieira.dev/gower-distance.html

Wahyuni, P. (2022). Antecedents and consequences of peer support. *International Journal of Applied Business and International Management, 7*(1), 75–95. https://doi.org/10.32535/ijabim.v7i1.1443

Zhang, J., Li, Y., Shen, F., He, Y., Tan, H., & He, Y. (2024). Hierarchical text classification with multi-label contrastive learning and KNN. *Neurocomputing*, *577*, 127323. https://doi.org/10.1016/j.neucom.2024.127323

ABOUT THE AUTHOR(S)

1st Author
Ko Chi Ling is a student at the School of Computing, INTI International College Penang, Malaysia. Her research focuses on software development, mobile applications, and the application of artificial intelligence in education, including projects on buddy systems with matching algorithms to enhance student support and peer mentoring. She has a strong interest in artificial intelligence, machine learning, and educational technology.
ORCID ID: https://orcid.org/0009-0001-2511-9361.
Email: p21013435@student.newinti.edu.my.

Journal of International Conference Proceedings (JICP)
Vol. 8 No. 1, pp.578-591, June, 2025
E-ISSN: 2621-993X P-ISSN: 2622-0989
https://ejournal.aibpmjournals.com/index.php/JICP/index

APPENDIX A

Figure A. Use Case Diagram

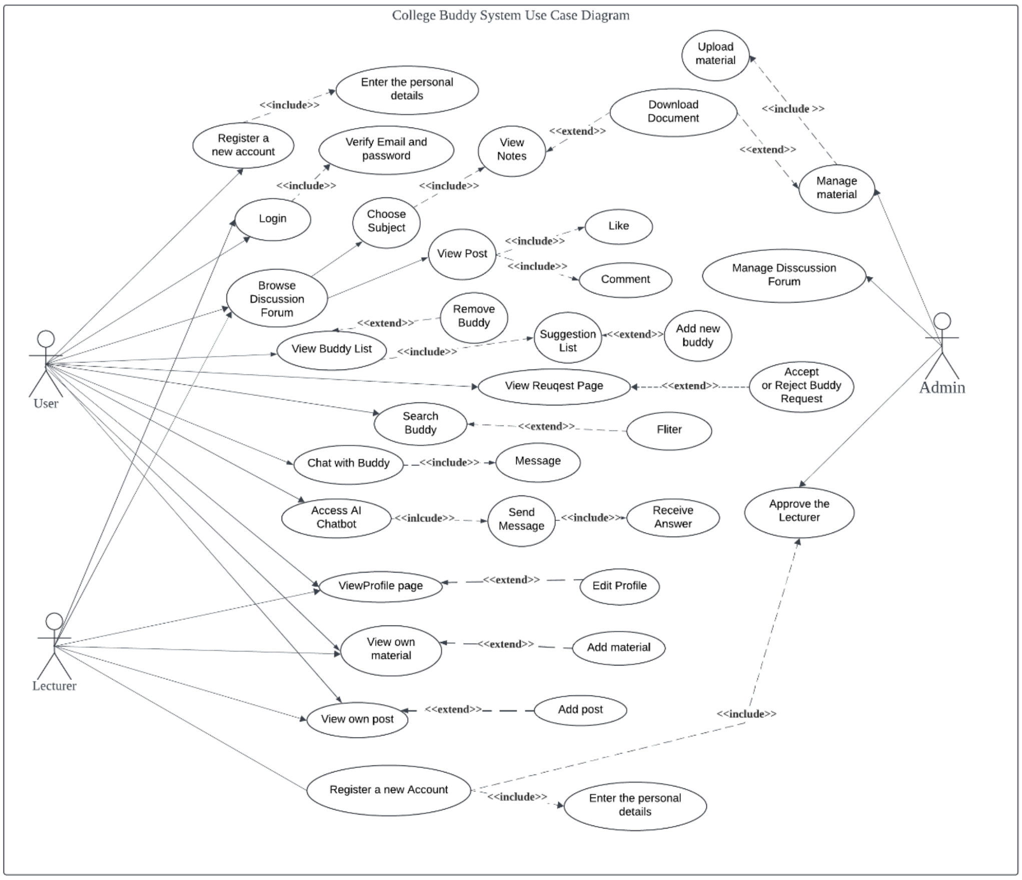


Figure B. Software Architecture Diagram