

УДК 004.8

Жук Д.І., Мазурець О.В., Кадинська В.Д., Тищенко О.О.

Хмельницький національний університет

ПІДХІД ДО ВИЗНАЧЕННЯ СУМІСНОСТІ КЛІЄНТІВ ШЛЮБНИХ АГЕНТСТВ ЗА ІНТЕЛЕКТУАЛЬНИМ АНАЛІЗОМ АНКЕТНИХ ДАНИХ

Запропоновано метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, який обчислює загальний відсоток сумісності пар, враховуючи вимоги до бажаних партнерів від обох сторін. Для реалізації запропонованого методу використовуються персональні дані клієнтів шлюбного агентства, взяті із знайденого датасету, який є у вільному доступі. На основі цих даних генеруються персональні текстові описи про клієнта і його вимоги до бажаного партнера, використовуючи існуючі нейромережі-трансформери. Для порівняння анкетних даних та обрахунку відсотку сумісності використовується архітектура сіамської нейронної мережі.

The method for determining the compatibility of clients of marriage agencies based on the intelligent analysis of questionnaire data is proposed, which calculates the overall percentage of compatibility of couples, taking into account the requirements for desired partners from both sides. To implement the proposed method, the personal data of the clients of the marriage agency, taken from the found dataset, which is freely available, is used. On the basis of this data, personal text descriptions about the client and his requirements for the desired partner are generated, using existing neural networks-transformers. Siamese neural network architecture is used to compare questionnaire data and calculate the compatibility percentage.

Сучасний темп життя, особливо у містах, спричиняє нестачу вільного часу, який можна було б присвятити побудові особистого життя, включаючи створення сім'ї. Надмірне навантаження на роботі, що забирає значну частину часу та особистих ресурсів, позбавляє людей можливості для знайомств, обмежуючи пошук потенційних майбутніх партнерів. У таких умовах звернення до послуг шлюбного агентства може стати ефективним рішенням, що дозволить знайти партнера, економлячи значну частину часу на пошуки.

Ще однією проблемою сучасності є підвищена вимогливість до партнерів, адже сучасне суспільство з кожним роком стає все більш вимогливим. Це пов'язано із значним розвитком та популяризацією психології, завдяки чому люди стають більш обізнані в тому як поводити себе в стосунках. Тому, людям значно важче знайти партнера, який буде відповідати всім поставленим якістьам.

Місце проживання, також, має великий вплив на процес пошуку партнерів, що особливо загострилось під час війни, коли багато людей були змушені виїхати за кордон [1]. В таких умовах, звернення до шлюбних агентств дає можливість врахувати географічне розташування при підборі партнерів.

Основна особливість діяльності шлюбних агентств полягає в ретельному підборі кандидатів для клієнтів, враховуючи їхні вимоги та побажання. Цей процес може охоплювати як традиційні індивідуальні консультації із спеціалістами, так і сучасні системи інтелектуального аналізу даних, що здатні визначати рівень сумісності на основі інформації з клієнтських анкет. Крім того, шлюбні агентства можуть надавати деякі додаткові послуги, такі як: психологічна підтримка, сімейна терапія та навчання по покращенню соціальних навичок.

Шлюбні агентства можуть різнитись за спеціалізацією та методами підбору партнерів, тому їх поділяють на три основні типи: традиційні шлюбні агентства, нішеві шлюбні агентства та шлюбні агентства онлайн [2].

Традиційні шлюбні агентства роблять акцент на особистому контакті з клієнтом. Вони зазвичай мають офіс, в який клієнти можуть звертатись для отримання послуг. Фахівці шлюбного агентства проводять з клієнтами індивідуальні консультації та створюють профіль клієнта, що слугуватиме основою для підбору потенційних партнерів. Додатково, клієнти можуть заповнювати анкети про себе, вказуючи особисту інформацію, таку як вік, стать та рід діяльності. Після збору інформації про клієнта, спеціалісти підбирають кандидата або декількох кандидатів та влаштовують їм особисті зустрічі, на яких люди мають змогу познайомитись та дізнатись більше один про одного, що допоможе їм визначити чи підходять вони одне одному [5]. Основною перевагою традиційного шлюбного агентства є те, що всі питання, пов'язані з підбором кандидатів і організацією зустрічей бере на себе агентство, що дозволяє клієнту зекономити багато часу. Однак, такі послуги можуть виявитись досить дорогими та довготривалими через обмежену клієнтську базу, яка може бути на момент звернення.

Нішеві шлюбні агентства зосереджені на пошуку партнерів для окремих груп. Наприклад, такі агентства можуть обслуговувати лише заможних клієнтів, які готові інвестувати великі суми для пошуку партнерів рівних за соціальним статусом. Також, нішеві шлюбні агентства можуть спеціалізуватися на підборі партнерів для клієнтів із певних релігійних груп, національностей чи культурних спільнот. Ці агентства орієнтовані на індивідуальні особливості групи людей із якими вони працюють, тому й звернення до нішевих шлюбних агентств може бути корисним для людей із певних соціокультурних груп.

Інший тип шлюбних агентств – це шлюбні агентства онлайн, які працюють виключно в інтернет-просторі. Вони можуть бути представлені у вигляді веб-сайтів або мобільних додатків, де клієнти самостійно реєструються та заповнюють анкетні дані. Пошук потенційних партнерів відбувається за допомогою спеціальних алгоритмів, які аналізують анкетні дані користувачів і пропонують їм список кандидатів із високим рівнем сумісності. Користувачі самі обирають з ким вони хотіли б познайомитись. Якщо, обоє з користувачів мають бажання дізнатись більше один про одного, вони можуть розпочати спілкування за допомогою вбудованих чатів.

Враховуючи соціальні проблеми сучасності, саме шлюбні агентства у онлайн форматі є найпопулярнішим та найефективнішим видом шлюбних агентств, через їхню простоту використання та доступність у будь-який час та незалежно від розташування. Крім цього, це досить економний варіант, адже користувачам подібних платформ доступний базовий безкоштовний варіант доступу до функціоналу та можливість придбання підписок різної вартості.

Отож, шлюбні агентства онлайн є найпопулярнішим та найефективнішим варіантом надання послуг з пошуку партнерів. Це зумовлено їхньою доступністю та простотою використання, адже звернутись до такого агентства можна з будь-якої точки світу. Ключовим етапом роботи шлюбних агентств є процес заповнення анкети клієнта, в яку людина додає як свою персональну інформацію, так і вимоги до бажаного партнера. Збереження анкетних даних користувачів у базі даних надає великий спектр можливостей для проведення інтелектуального аналізу збережених анкет, що є важливим при автоматизації процесів підбору партнерів. Тому, заповнення анкет та їхня подальша обробка є основою для реалізації методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних.

Метою роботи є розробка методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, та його прикладна реалізація у вигляді веб-сайту для пошуку партнера з високою сумісністю.

Для автоматизованого визначення сумісності клієнтів шлюбних агентств за їхніми анкетними даними необхідно використовувати сучасні методи інтелектуального аналізу даних. Інтелектуальний аналіз даних (ІАД) являє собою процес виявлення прихованих, невідомих або неочевидних закономірностей у великих обсягах даних на основі використання математичних моделей, статистичних методів чи алгоритмів штучного інтелекту [3, 4]. Основною метою інтелектуального аналізу даних є перетворення необроблених даних на інформацію, яка може бути використана для прийняття рішень, прогнозування чи оптимізаційних процесів [5].

Обраний метод потрібно адаптувати під вирішення конкретної задачі та навчити на відповідному наборі навчальних даних [6, 7]. Під час навчання метод аналізує зв'язки між вхідними даними та навчається знаходити закономірності [8].

Після проведення навчання необхідно провести оцінку точності роботи методу. Метод оцінюється на даних, які не були використанні під час навчання, що дозволяє оцінити якість роботи методу з незнайомою інформацією [9]. Для цього, також, використовуються різні метрики, вибір яких залежить від обраного методу [11].

Якщо навчання методу досягло задовільного рівня то можна переходити до наступного етапу, який полягає в інтерпретації та візуалізації отриманих результатів. Для цього створюються різноманітні графіки та діаграми, які допомагають краще зрозуміти отримані результати та прийняти подальше рішення щодо використання натренованого методу [12].

Метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних складається із трьох окремих етапів (Рисунок 1). На першому етапі формуються анкети клієнтів на основі персональних даних із датасету. Другий етап передбачає підготовку анкетних даних для проведення інтелектуального аналізу. Останній третій етап присвячений обрахунку відсотків сумісності клієнтів шлюбних агентств на основі їхніх анкетних даних.



Рисунок 1 – Схема методу визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних

Для реалізації запропонованого методу використовуються персональні дані клієнтів шлюбного агентства, взяті із знайденого датасету, який є у вільному доступі. На основі цих даних генеруються персональні текстові описи про клієнта і його вимоги до бажаного партнера, використовуючи промпти та ChatGPT [20]. Для порівняння анкетних даних та обрахунку відсотку сумісності використовується архітектура сіамської нейронної мережі.

Отже, на першому етапі методу відбувається створення анкет клієнтів шлюбного агентства, використовуючи дані із датасету. Оскільки, обраний датасет містить лише персональні та категоріальні дані, такі як рівень освіти, список хобі чи знак зодіаку, було вирішено доповнити анкети згенерованими текстовими описами користувачів про себе. Також, кожна анкета буде доповнена інформацією про вимоги до бажаного партнера. Ця інформація буде обрана випадковим чином із числа доступних даних, що зберігатимуться у базі даних. На основі сформованих вимог, також, буде згенеровано текстовий опис бажаного партнера.

Оскільки, згенерований текст може не відповідати очікуванням клієнта, перед проведенням пошуку партнера, користувачеві веб застосунку буде доступна можливість обрати підходящі текстові описи з числа згенерованих чи відредагувати їх, або згенерувати нові текстові описи. Це допоможе уникнути аналізування погано згенерованих текстів, які можуть погіршувати результати пошуку бажаних партнерів.

Для проведення інтелектуального аналізу анкетних даних необхідно провести попередню обробку даних, щоб уникнути отримання некоректних результатів через наявність помилкових вхідних даних. Тому, всі підготовчі кроки відбуватимуться на другому етапі. Спершу, необхідно провести перевірку на наявність пустих полів в анкетах клієнтів. При відсутності даних їх буде додано випадковим чином із числа доступних варіантів вибору. Наприклад, якщо в користувача відсутні дані про його вік, йому буде присвоєно випадкове число в діапазоні від 18 до 122 років. За нижню межу допустимого віку прийнято вік досягнення повноліття [21], а за верхню – вік найстаршої людини в світі [22].

Оскільки, крім персональних та категоріальних даних у анкетах присутні текстові описи за якими будуть обраховуватись відсотки подібності, виникає необхідність провести попередню обробку текстів перед початком аналізу. Для цього буде застосовано нормалізацію текстового вмісту. Під нормалізацією тексту мається на увазі приведення текстового вмісту до єдиної форми, яка використовуватиметься для подальшого аналізу. Зазвичай, це включає переведення тексту в нижній регістр [23]. Це допомагає уникнути ситуації, коли одне і те саме слово сприймається програмою по різному, через різні варіанти написання. Також, даний крок включатиме видалення зайвих символів та знаків із тексту, адже в даному контексті вони не несуть жодного семантичного навантаження.

Далі, текстові описи необхідно векторизувати, адже векторизація тексту – це основний процес, який необхідний для обробки природної мови. В

загальному, векторизація – це представлення текстового вмісту у числовий формат, зрозумілий для алгоритмів [24]. Саме на основі векторизованого тексту, сучасні системи для обробки природної мови здатні вирішувати безліч завдань, починаючи від пошуку ключових слів і закінчуючи розумінням ідей із тексту.

На останньому третьому етапі методу відбувається обрахунок відсотків сумісності клієнтів шлюбного агентства за інтелектуальним аналізом анкетних даних. Для початку, формується вибірка потенційних партнерів, які відповідають зазначеним вимогам. Вибірка формується шляхом порівняння вимог до партнера із анкетними даними клієнтів зареєстрованих у базі даних. Для цього необхідно створити певні правила, за якими буде відбуватись фільтрація. Наприклад, якщо клієнт шукає партнера віком від 20 до 25 років то в вибірку потраплять лише ті люди, чий вік задовольнятиме вказану вимогу.

Далі, відбувається обрахунок оцінок подібності текстових описів клієнтів, використовуючи архітектуру сіамської нейронної мережі. Для цього відбувається перехресне порівняння текстових описів для кожної потенційної пари. Тобто, опис бажаного партнера для цільового клієнта порівнюється із текстовими розповідями потенційних партнерів про себе для отримання відсотку схожості текстових описів. Потім, відбувається обернене порівняння, де розповідь про себе від цільового клієнта порівнюється із описами бажаних партнерів від потенційних партнерів із вибірки. Це необхідно для того, щоб враховувати сумісність відносно вимог обох партнерів.

Після того, як буде обчислено дві оцінки схожості текстових описів, обчислюватиметься загальна оцінка сумісності, шляхом обрахунку середнього арифметичного отриманих чисел. Таким чином, буде отримано загальний відсоток сумісності для кожної аналізованої пари анкет, який можна відсортувати та відобразити користувачеві у вигляді переліку із п'яти партнерів із найбільшим відсотком сумісності.

Отже, було розроблено метод визначення сумісності клієнтів шлюбних агентств за інтелектуальним аналізом анкетних даних, який обчислює загальний відсоток сумісності пар, враховуючи вимоги до бажаних партнерів від обох сторін.

Перелік посилань

1. Потапчук Є. М., Балукова А. І. Чинники міжособистісної сумісності майбутніх шлюбних партнерів. Причорноморський науково-дослідний інститут економіки та інновацій. Одеса, 2021. С167–171.
2. Потапчук Н., Чорноморець Д. Життєдіяльність сімей військовослужбовців в умовах війни як психологічна проблема. *Psychology Travelogs*, 2024. С 241–250.
3. Мазурець О.В., Віт Р.В. Дослідження ефективності методу виявлення цільових об'єктів предметної області. Інформаційні технології і автоматизація. Матеріали XVII міжнародної науково-практичної конференції. 31 жовтня – 1 листопада 2024 р. Одеса, ОНТУ. 2024. С.650-653.
4. Mazurets O., Tymofiiiev I., Dydo R. Approach for Using Neural Network BERT-GPT2 Dual Transformer Architecture for Detecting Persons Depressive State. *Ricerche scientifiche e metodi*

- della loro realizzazione: esperienza mondiale e realtà domestiche. Raccolta di articoli scientifici con gli atti della VI Conferenza scientifica e pratica internazionale. 15 novembre, 2024. Bologna, Repubblica Italiana. 2024. Pp. 147-151.
5. Blazhuk V., Mazurets O., Zalutskaja O. An Approach to Using the mBERT Deep Learning Neural Network Model for Identifying Emotional Components and Communication Intentions. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. October 23-25, 2024. Dubrovnik, Croatia. 2024. Pp. 79-84.
 6. Мазурець О.В., Молчанова М.О., Кліменко В.І., Собко О.В., Супрун П.К. Даталогічна модель бази даних для виявлення гендерної приналежності за SVM-аналізом дописів інтернет-мереж з використанням об'єктно-орієнтованого проєктування. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №3, Т.2 (337). С. 197-204.
 7. Mazurets O., Molchanova M., Klimenko V., Prosvitliuk M Practice Implementation of Neural Network Model BART-Large-CNN for Text Annotation. Prospects of Scientific Research in the Conditions of the Modern World. Proceedings of XXVII International scientific and practical conference. June 12-14, 2024. Rotterdam, Netherlands. 2024. Pp. 97-102.
 8. Sobko O., Mazurets O., Didur V., Chervonchuk I. Recurrent Neural Network Model Architecture for Detecting a Tendency to Atypical Behavior Of Individuals by Text Posts. Theoretical and Practical Aspects of Modern Research. Proceedings of XXVI International scientific and practical conference. June 5-7, 2024. International Scientific Unity. Ottawa, Canada. 2024. Pp. 113-117.
 9. Mazurets O.V., Sobko O.V., Molchanova M.O., Zalutskaja O.O., Yurchak A.V. Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts. Global Science: Prospects and Innovations. Proceedings of the II International Scientific and Theoretical Conference «Scientific Review of the Actual Events, Achievements and Problems». May 31, 2024. Berlin, Federal Republic of Germany: International Center of Scientific Research. 2024. Pp. 160-167.
 10. Molchanova M., Mazurets O., Sobko O., Boiarchuk I. Object-Oriented Approach for Ethnic Enmity Detection in Text Messages by NLP. Proceedings of XXI International Scientific and Practical Conference «Scientific Achievements and Innovations as a Way to Success». May 1-3, 2024. Vilnius, Lithuania. 2024. Pp. 73-77.
 11. Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №2 (333). С. 200-206.
 12. Залуцька О.О., Молчанова М.О., Віт Р.В., Мазурець О.В. Конфігурування нейронної мережі для класифікації емоційної тональності текстової інформації за показниками семантичної зв'язності. Збірник наукових праць за матеріалами XV Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2023». Хмельницький, 2023. с. 102-107.