

Системный подход и методы машинного обучения в прогнозировании психоэмоциональных состояний по цифровой активности в социальных сетях

Ю.С. Романова, Е.В. Пастухова

*Санкт-Петербургский государственный университет аэрокосмического
приборостроения, г. Санкт-Петербург*

*Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В. И. Ульянова (Ленина), г. Санкт-Петербург*

Аннотация: Статья посвящена системному анализу влияния цифровой активности в социальных сетях на психоэмоциональные показатели: тревожность, депрессию, СДВГ и самооценку. На основе анкетных данных пользователей реализованы этапы предобработки, построения корреляционных зависимостей и кластеризации методом k-средних. Основное внимание уделено разработке и обучению ансамблевой мета-модели, сочетающей логистическую регрессию, наивный байесовский классификатор и случайный лес. В результате анализа выявлены группы пользователей, наиболее подверженные негативному влиянию социальных сетей. Установлены статистически значимые взаимосвязи с проблемами сна, рассеянностью и тревожностью. Сформулированы рекомендации по поддержанию психического здоровья для различных возрастных групп.

Ключевые слова: системный анализ, цифровая активность, социальные сети, машинное обучение, кластеризация, корреляционный анализ, цифровая зависимость, психоэмоциональное состояние, интеллектуальный анализ информации.

В условиях стремительного развития цифровых технологий и роста популярности социальных сетей наблюдается существенное изменение структуры повседневной коммуникации, особенно среди молодежи. Эти изменения сопровождаются не только новыми возможностями для социального взаимодействия, но и рядом потенциальных рисков для психического здоровья пользователей. Участвовавшие случаи тревожных расстройств, сниженной самооценки, нарушений сна и симптомов, схожих с

синдромом дефицита внимания с гиперактивностью (СДВГ), всё чаще связываются с чрезмерным использованием цифровых платформ. В связи с этим, в последние годы всё большее число исследований посвящается использованию методов машинного обучения для анализа психоэмоционального состояния на основе цифрового поведения пользователей в социальных сетях. Как показано в библиометрическом исследовании [1], наблюдается устойчивый рост интереса к данной теме со стороны научного сообщества, особенно в контексте выявления тревожности, депрессии и других расстройств через анализ цифрового следа.

Актуальность настоящего исследования обусловлена необходимостью выявления устойчивых закономерностей между характером цифровой активности и психоэмоциональными нарушениями с целью их последующего анализа и интерпретации. Системный анализ, как междисциплинарный подход, позволяет рассматривать проблему комплексно, охватывая как информационно-поведенческие, так и психофизиологические аспекты. Его применение обеспечивает построение целостной модели, в которой цифровое поведение рассматривается как элемент сложной информационной системы, взаимодействующей с внутренними психоэмоциональными процессами личности.

Применение методов интеллектуального анализа данных и искусственных нейронных сетей в смежных областях, например, при изучении зависимости от мобильных игр с учётом поведенческих и социокультурных факторов [2], демонстрирует высокий потенциал подобных подходов и в контексте анализа цифровой активности в социальных сетях.

Цель исследования - проведение системного анализа влияния цифровой активности в социальных сетях на психоэмоциональные показатели с применением методов машинного обучения для выявления закономерностей

и групп риска. Для достижения этой цели были поставлены следующие задачи:

- собрать и структурировать анкетные данные, включающие показатели цифровой активности и психического здоровья пользователей;
- выполнить предобработку данных (устранение выбросов, заполнение пропущенных значений, нормализацию данных);
- провести корреляционный анализ и визуализировать взаимосвязи между переменными;
- осуществить кластеризацию данных методом k-средних для выделения групп пользователей по психоэмоциональным признакам;
- разработать и обучить ансамблевую мета-модель машинного обучения для выявления скрытых закономерностей;
- проанализировать полученные результаты и выявить группы риска;
- сформулировать рекомендации по снижению негативного влияния социальных сетей на психическое здоровье.

Материалы и методы

В настоящем исследовании используются методы машинного обучения, обладающие способностью выявлять скрытые взаимосвязи в больших поведенческих массивах. Как отмечается в [3], применение технологий искусственного интеллекта в задачах анализа данных способствует более точной и своевременной поддержке принятия решений в условиях высокой сложности и неопределённости.

Для программной реализации методики исследования была выбрана технология Python, так как она широко используется в области машинного обучения и анализа данных [4].

Анкетирование участников исследования было проведено с использованием платформы дистанционного обучения LMS Санкт-Петербургского государственного университета аэрокосмического

приборостроения (ГУАП). В исследовании приняли участие 473 респондента, из которых 264 человека составили женщины и 209 человек - мужчины. Возраст 73% участников не превышал 35 лет. Социальная структура выборки распределилась следующим образом: 68% — студенты, 27% — преподаватели, 5% — представители учебно-вспомогательного персонала. Анкета включала вопросы о времени, проводимом в социальных сетях, а также шкалы для оценки уровня тревожности, депрессии, признаков СДВГ и самооценку. При обработке информации соблюдены требования конфиденциальности и сохранности персональных данных в соответствии с действующим законодательством.

Результаты исследования

Подход к выбору признаков цифровой активности и психоэмоциональных показателей основывался на обобщённых результатах систематического обзора [5], в котором были выделены ключевые маркеры депрессивных и тревожных состояний, проявляющиеся в поведении пользователей социальных сетей. Тестирование, использованное в исследовании, включало как социально-демографические, так и психоэмоциональные блоки вопросов. Участникам предлагалось указать возраст, пол, профессию. Далее следовали вопросы, связанные с характером использования социальных сетей: предпочитаемые платформы, среднее время использования, частота бесцельного просмотра и отвлечений во время других занятий, а также наличие тревожности при временном отказе от цифровой активности. Часть вопросов использовала шкальные оценки, что позволило впоследствии применять количественные методы анализа.

Оценка частоты и интенсивности проявлений психоэмоциональных состояний осуществлялась посредством 12 вопросов, представленных в формате шкалы Лайкерта с градацией от 0 до 5 баллов [6]. Нулевая оценка соответствовала отсутствию или крайне низкой выраженности симптома,

тогда как значение 5 отражало его максимальную выраженность или частоту. Данный подход позволил количественно зафиксировать индивидуальные различия в проявлениях психических состояний, ассоциированных с цифровой активностью. Показатели включали следующие аспекты: бесцельное использование социальных сетей, отвлечённость во время выполнения задач и легкость отвлечения (связанные с признаками СДВГ); тревожность, возникающая при отсутствии доступа к социальным сетям, наличие беспокойства и трудности концентрации (тревожные проявления); склонность к социальным сравнениям, эмоциональная реакция на такие сравнения, потребность в подтверждении через цифровые механизмы (компоненты самооценки); чувство подавленности, колебания интереса к повседневной активности и проблемы со сном (депрессивные симптомы).

Для интерпретации совокупной оценки психоэмоционального состояния был использован обоснованный пороговый подход. При условии, что респондент набирает по 3 балла из 5 по каждому из 12 шкальных вопросов, итоговый балл составил 36, что отражает наличие лёгких или умеренных симптомов по всем измеряемым аспектам психического здоровья. Однако такая выраженность признаков может и не свидетельствовать о клинически значимых нарушениях. В связи с этим, на основе экспертного обоснования и эмпирического распределения данных, было установлено рациональное пороговое значение в 40 баллов. Превышение этого значения интерпретируется как наличие устойчивых и выраженных психоэмоциональных симптомов, что с высокой вероятностью указывает на необходимость проведения углублённой психологической диагностики. При анализе анкет выяснилось, что приблизительно 37% из 473 участников исследования набрали 40 баллов и более, что указывает на серьезные признаки психических проблем у респондентов.

Собранные данные были предварительно обработаны путем выявления пропущенных значений, устранения выбросов и нормализации [7]. После этого осуществлен анализ корреляции между признаками.

Из приведенной ниже корреляционной таблицы (рис. 1) можно сделать вывод, что время, проведенное на различных платформах социальных сетей, имеет умеренную положительную корреляцию с СДВГ, тревогой и депрессией со значениями коэффициента корреляции (r) 0,45, 0,44 и 0,35 соответственно; корреляция с показателями самооценки находится на положительной и слабой стороне со значением 0,138.

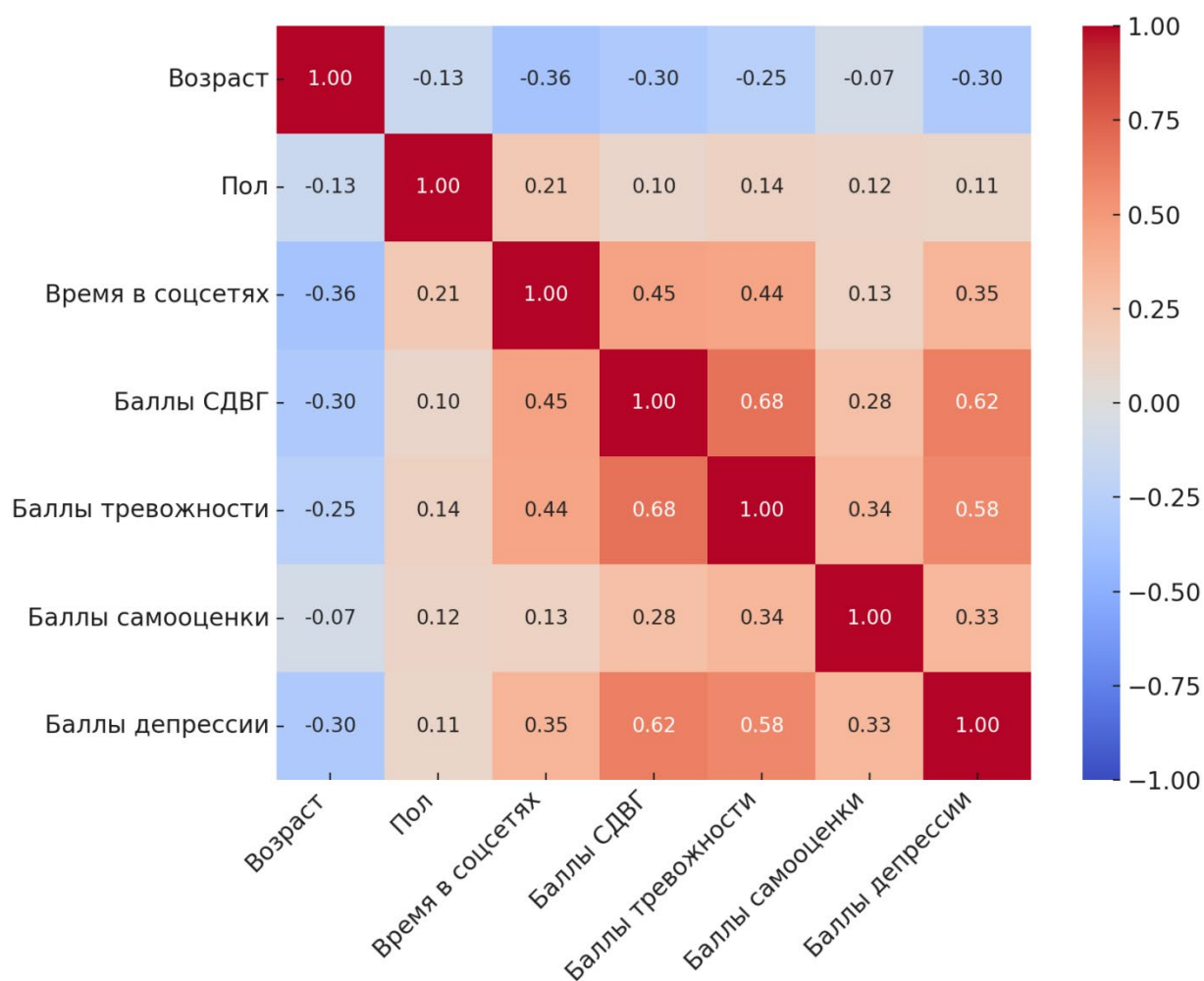


Рис 1. Корреляционная таблица

Существует отрицательная слабая корреляция между возрастом и всеми другими переменными. Интерпретация этого факта может заключаться в том,

что чем выше возраст участника, тем ниже будут его показатели использования социальных сетей и психического здоровья. Отметим, что это слабая корреляция со значениями от -0,36 до -0,07 для всех переменных.

В анализируемом наборе данных наблюдается мультиколлинеарность низкой и средней степени между некоторыми независимыми признаками. Значения корреляции в пределах $0,11 < r < 0,68$ свидетельствуют о наличии взаимосвязей, однако они не критичны и не нарушают устойчивость модели при условии применения регуляризации или отбора признаков. Это означает, что большинство переменных сохраняют достаточную независимость, позволяя использовать их совместно в алгоритмах машинного обучения без существенного риска переобучения или искажения весов. Как отмечается в [8], существует вероятность того, что алгоритмы машинного обучения смогут обнаружить скрытые закономерности на этапе обучения прогнозирующей модели.

На следующем этапе исследования данные были разбиты на кластеры по выбранным признакам. Результаты кластеризации позволили выделить три отчетливо различающиеся группы пользователей по характеру использования социальных сетей и выраженности психоэмоциональных состояний (таблица №1).

Таблица №1

Кластеризация данных

Показатель	Кластер 1	Кластер 2	Кластер 3
Средний возраст	23,68	31,52	24,75
Среднее время в соцсетях (часы)	5,2	3,15	4,15
Отвлекаются на соцсети (раз в день)	4,3	2,26	3,16
Беспокойство при отсутствии доступа (раз в день)	3,72	1,69	2,19
Частота подавленности (оценка от 1 до 5)	3,96	1,9	3,53
Проблемы со сном (оценка от 1 до 5)	3,68	1,98	3,58

Кластер 1 включает преимущественно молодых пользователей с самой высокой цифровой активностью. Участники этой группы часто отвлекаются на социальные сети и испытывают тревожность при отсутствии доступа. Также именно в этом кластере наблюдаются наиболее выраженные признаки подавленности и проблем со сном. Эти данные позволяют охарактеризовать кластер как «высокорисковую группу», потенциально уязвимую к негативным психологическим последствиям цифровой активности. Кластер 2 характеризуется наименьшими значениями по всем показателям. Пользователи старше, проводят меньше времени в соцсетях, реже отвлекаются и не испытывают значимого уровня тревожности, подавленности или проблем со сном. Этот кластер можно интерпретировать как «наиболее благополучную группу», демонстрирующую устойчивое поведение и невысокий уровень цифровой зависимости. Кластер 3 занимает промежуточное положение. Возраст участников близок к кластеру 1, как и уровень вовлеченности в соцсети, но значения показателей психического здоровья несколько лучше, чем в кластере 1. Это может свидетельствовать о «начальной стадии формирования зависимости», либо об индивидуальных защитных механизмах участников.

Таким образом, кластерный анализ позволил не только сегментировать пользователей, но и выявить группы риска, нуждающиеся в профилактике цифровой зависимости и поддержке психоэмоционального благополучия. Результаты кластеризации были использованы на завершающем этапе исследования при построении модели машинного обучения.

На основании предварительных экспериментальных испытаний, включающих комбинирование различных базовых моделей машинного обучения и оценку их прогностической точности, в настоящем исследовании была выбрана ансамблевая мета-модель. Её архитектура построена на объединении логистической регрессии и гауссовского наивного байесовского

классификатора в качестве базовых моделей, при этом роль оценщика выполняет алгоритм случайного леса. Данная комбинация обеспечивает баланс между интерпретируемостью (за счёт логистической регрессии), устойчивостью к шуму и малым выборкам (свойственной наивному байесовскому классификатору) и высокой обобщающей способностью (обеспечиваемой случайным лесом). Включение случайного леса в ансамблевую мета-модель позволяет обеспечить устойчивость к переобучению, высокую точность классификации и интерпретируемость результатов. Кроме того, возможна формализация структуры модели средствами сетей Петри, что открывает перспективы для дальнейшей верификации и оптимизации архитектуры алгоритма [9]. Кроме того, использование ансамблевого подхода позволяет нивелировать слабости отдельных моделей и повысить общую точность, устойчивость и стабильность прогноза, особенно в условиях слабовыраженной линейной структуры и потенциальной мультиколлинеарности между признаками. Как отмечается в [9], несмотря на популярность методов глубокого обучения, традиционные ML-модели остаются предпочтительными в задачах анализа психоэмоционального состояния в условиях ограниченности данных и необходимости прозрачности принятия решений. Предложенная модель направлена не только на прогнозирование вероятности наличия у пользователя выраженных психоэмоциональных симптомов, требующих профессионального вмешательства, но и на автоматическое отнесение его к одной из ранее выделенных кластерных групп, что позволяет учитывать индивидуальные особенности поведения и уровня цифровой активности при последующем анализе и интерпретации данных.

Заключение

В ходе настоящего исследования был выполнен системный анализ влияния цифровой активности в социальных сетях на психоэмоциональные

показатели пользователей с применением современных методов машинного обучения. На основе анкетирования 473 респондентов, охватывающего различные аспекты цифрового поведения и психического состояния, были реализованы этапы предобработки данных, корреляционного анализа, кластеризации методом k-средних и построения прогностической мета-модели.

Результаты анализа подтвердили наличие статистически значимых взаимосвязей между временем, проводимым в социальных сетях, и такими показателями, как рассеянность, тревожность, снижение самооценки, подавленность и проблемы со сном. Кластеризация позволила выделить группы пользователей с различным уровнем вовлечённости и психоэмоциональной уязвимости. Разработанная ансамблевая мета-модель эффективно классифицирует пользователей по степени риска, учитывая индивидуальные особенности цифрового поведения.

На основе полученных данных и, опираясь на исследования [10,11], могут быть сформулированы следующие рекомендации, направленные на поддержание психического здоровья пользователей:

- ограничивать бесцельное использование социальных сетей и внедрять цифровые перерывы в течение дня;
- отслеживать и осознавать эмоциональные реакции, возникающие при сравнении себя с другими пользователями;
- развивать навыки саморегуляции и концентрации внимания, особенно среди молодёжи.

Полученные результаты подтверждают актуальность системного подхода к анализу цифрового поведения и демонстрируют возможности интеллектуальных моделей как инструмента раннего выявления цифровой зависимости и психоэмоциональных рисков.

Литература

1. Kim J, Lee D, Park E. Machine Learning for Mental Health in Social Media: Bibliometric Study // Journal of Medical Internet Research, 2021, №23 (3). P. 24870. URL: doi 10.2196/24870
2. Хозяшева А.А. Потенциал нейронных сетей для обнаружения зависимости от мобильных игр: проверка концепции в российском контексте // Инженерный вестник Дона, 2025, №5. URL: ivdon.ru/ru/magazine/archive/n5y2025/10030
3. Тимохин М.Ю., Шаранин В.Ю. Искусственный интеллект и теория принятия решений: современные тенденции // Инженерный вестник Дона, 2023, №10. URL: ivdon.ru/ru/magazine/archive/n10y2023/8746
4. Маккинни Y. Python и анализ данных. Москва: ДМК-Пресс, 2023. 536с.
5. Salas-Zárate R, Alor-Hernández G, et al. Detecting Depression Signs on Social Media: A Systematic Literature Review // Healthcare (Basel), 2022, №10(2). P., 291. URL: doi 10.3390/healthcare10020291
6. Квон Г. М., Вакс В. Б., Поздеева О. Г. Использование шкалы Лайкерта при исследовании мотивационных факторов обучающихся // Научно-методический электронный журнал «Концепт», 2018, № 11. С. 84–96.
7. Романова Ю.С., Пастухова Е.В. Эффективные подходы к подготовке данных для машинного обучения // XXVII МНК «Волновая электроника и инфокоммуникационные системы». СПб: ГУАП, 2024. С.87-91.
8. Николенко С. Машинное обучение. Основы. СПб: Питер, 2025. 608с.
9. Петросов Д.А. Моделирование алгоритма машинного обучения «Случайный лес» математическим аппаратом теории сетей Петри // Инженерный вестник Дона, 2024, №10. URL: ivdon.ru/ru/magazine/archive/n10y2024/9561

10. Ding, Z., Wang, Z., Zhang, Y. et al. Trade-offs between machine learning and deep learning for mental illness detection on social media//Scientific Reports, 2025, №15. P. 14497. URL: doi.org/10.1038/s41598-025-99167-6
11. Sandra P. Thomas, Youth A. Digital Devices and Advice for Parents, Issues in Mental Health//Nursing, 2025, №46 (3). URL: 10.1080/01612840.2025.2466997

References

1. Kim J, Lee D, Park E. Journal of Medical Internet Reseach, 2021, №23 (3). P. 24870. URL: doi 10.2196/24870
2. Khozyasheva A.A. Inzhenernyj vestnik Dona, 2025, №5. URL: ivdon.ru/ru/magazine/archive/n5y2025/10030
3. Timokhin M.Yu., Sharanin V.Yu. Inzhenernyj vestnik Dona, 2023, №10. URL: ivdon.ru/ru/magazine/archive/n10y2023/8746
4. Makkinni Y. Python i analiz dannykh [Python and Data Analysis]. Moskva: DMK-Press, 2023. 536p.
5. Salas-Zárate R, Alor-Hernández G, et al. Healthcare (Basel), 2022, №10 (2). P. 291. URL: doi 10.3390/healthcare10020291
6. Kvon G. M., Vaks V. B., Pozdeeva O. G. Nauchno-metodicheskiy elektronnyj zhurnal «Kontsept», 2018. № 11. pp. 84–96.
7. Romanova Yu.S., Pastukhova E.V. XXVII MNK "Volnovaya elektronika i infokommunikatsionnye sistemy". SPb: GUAP, 2024. pp.87-91.
8. Nikolenko S. Mashinnoe obuchenie. Osnovy [Machine Learning. Basics]. SPb: Piter, 2025. 608p.
9. Petrosov D.A. Inzhenernyj vestnik Dona, 2024, №10. URL: ivdon.ru/ru/magazine/archive/n10y2024/9561
10. Ding, Z., Wang, Z., et al. Trade-offs between machine learning and deep learning for mental illness detection on social media. Scientific Reports, 15. P. 14497. URL: doi.org/10.1038/s41598-025-99167-6



11. Sandra P. Thomas, Youth Anxiety. Nursing, 2025, №46(3). URL:
10.1080/01612840.2025.2466997.

Дата поступления: 27.06.2025

Дата публикации: 25.08.2025