

Определение критических факторов для дородового наблюдения с помощью двухфазного машинного обучения

Игорь А. Бессмертный

Университет ИТМО

bessmertny@itmo.ru

Бидру Ч. Енкомирам

Университет ИТМО

checkt.birsh@gmail.com

Адане Л. Мамуйе

Университет Гондар

adane.letta@uog.edu.et

Аннотация. Определение критических факторов, влияющих на дородовые визиты беременных, позволит внести большой вклад в снижение материнской и младенческой смертности. В данном исследовании предпринята попытка построить кластерную прогностическую модель для определения критических факторов дородового наблюдения среди женщин в Эфиопии с использованием двухэтапных методов машинного обучения. Использовались 5990 предварительно обработанных акушерских данных с 13 выбранными атрибутами, которые были собраны в исследовательском центре Dabat при Университете Гондара, Эфиопия. Для решения поставленной задачи были разработаны шесть кластерных моделей с использованием алгоритмов K-means, Expectation Maximization и MakeDensity алгоритмы кластеризации, которые хорошо подходят для анализа наборов данных о здоровье, обеспечивая масштабируемость, гибкость, устойчивость к шуму и способность обрабатывать неравномерные размеры кластеров. Алгоритм K-means превзошел алгоритмы кластеризации MakeDensity и Expectation Maximization с минимальным количеством квадратов ошибок (17094) и количеством итераций (три). Затем результаты кластерной модели K-means были использованы в качестве входных данных для сегментации набора данных и построения 21 прогностической модели с использованием древовидного классификатора J48, генераторов правил PART и JRip при настройке их параметров. Модель для прогнозирования последующего дородового наблюдения построена путем индукции усеченного правила PART (точность 99,93%) и извлекает интересные атрибуты, такие как посещение женской консультации, местоположение женской консультации, частота посещения, текущая вакцинация, возраст женщины и тестирование на ВИЧ, чтобы предсказать последующее наблюдение в женской консультации в районе исследования.

Ключевые слова: машинное обучение; материнское здоровье; акушерство; K-means; алгоритм PART; дородовое наблюдение; Эфиопия

I. ВВЕДЕНИЕ

Младенческая и материнская смертность в странах Азии и Африки, обусловленная низким охватом дородовым наблюдением, является серьезной проблемой. [1]. Дородовое наблюдение (АНК), наряду с планированием семьи, квалифицированным родовспоможением и неотложной акушерской помощью, является ключевым элементом пакета услуг,

направленных на улучшение здоровья матери и новорожденного [2, 3]. Согласно рекомендациям Всемирной организации здравоохранения (ВОЗ), каждая беременная женщина должна получить не менее четырех посещений АНК во время беременности [4]. Однако в Эфиопии беременные женщины часто приступают к АНК только после первого триместра и не достигают рекомендованного количества посещений АНК [5]. В результате в 2017 году материнская смертность оценивается в 401/100 000, что означает, что около 12 000 матерей умирают каждый год, а неонатальная смертность составляет 29/1000 живорожденных, что связано с прямыми акушерскими осложнениями, составляющими 85 % смертности и заболеваемости матерей [6].

Определение критических факторов, влияющих на дородовые визиты, внесет большой вклад в снижение материнской и младенческой смертности. Однако до сих пор для выявления критических факторов использовались только статистические методы или методы одношагового машинного обучения (ML). Статистические инструменты имеют ограниченные возможности, и автоматически добывать в них ценную информацию достаточно сложно [7].

В этом исследовании мы сначала применили методы обучения без учителя для получения результатов кластеризации. Затем лучший результат кластеризации использован в качестве входных данных для методов обучения с учителем, поскольку исходный набор данных, используемый для этого исследования, содержит неразмеченные данные.

Таким образом, основной целью исследования является разработка модели, предсказывающей наблюдение за беременными с помощью двухфазных методов машинного обучения.

II. СОСТОЯНИЕ ПРОБЛЕМЫ И ТЕКУЩИЕ ИССЛЕДОВАНИЯ

Предыдущие исследования, такие как [8, 9, 10, 11], были проведены для определения критических факторов, способствующих материнской и детской смертности с использованием статистического пакета для социальных наук (SPSS) в качестве средства исследования.

Шахджахан и др. [8] оценили факторы, связанные с практикой охраны материнства в сельских районах Бангладеш, и изучили связь между использованием услуг дородового ухода и социально-демографическими параметрами сельских женщин. Берекет [9] исследовал

использование матерями дородового ухода и их удовлетворенность услугами родовспоможения в отдельных государственных медицинских учреждениях зоны Волайта, Южная Эфиопия.

Кассу [10] изучил определяющие факторы, влияющие на использование услуг по охране материнского здоровья в сельских районах Эфиопии, на основе данных Эфиопского исследования демографии и здоровья (EDHS) 2011 года. Теводрос [11] провел еще одно исследование, посвященное использованию матерями услуг дородового ухода среди подростков в Эфиопии.

Тарику [12] разработал прогностическую модель предпочтений в отношении рождаемости для женщин репродуктивного возраста с использованием методов интеллектуального анализа данных. Теклехайманот [13] определил факторы, предсказывающие наблюдение женщин за дородовым наблюдением во время беременности в Эфиопии с помощью технологии классификации данных на основе набора данных EDHS 2011. Теклехайманот [13] попытался показать применимость J48, наивного Байеса и нейронной сети с уже имеющимися. Усеченный набор J48 [13] с выбранным атрибутом на 80-процентном разбиении показал самую высокую точность классификации (79,85%).

В настоящем исследовании для моделирования кластера использовались методы K-means, Expectation maximisation (EM) и Make-Density clustering, а для выявления факторов продолжения дородового наблюдения – алгоритмы J48, PART и JRip на основе местных данных системы медико-демографического наблюдения Dabat (Dabat HDSS) при университете Гондара, Эфиопия.

III. МЕТОДЫ ИССЛЕДОВАНИЯ

A. Источник данных

Данные, использованные в исследовании, были собраны из Dabat HDSS в Университете Гондара, Эфиопия, которые охватывают период с 2009 по май 2017 года, содержат более 6700 экземпляров и 40 атрибутов [17]. Dabat HDSS был создан для получения данных о здоровье и населении на уровне района и обеспечения базы исследования и выборки для исследований на уровне общин [18].

B. Подготовка данных

Чтобы подготовить данные к анализу и построить соответствующие кластерные и прогностические модели, выполняются задачи предобработки. При очистке данных, заполнении недостающих значений, обработке зашумленных и отброшенных данных использовался режим EM. Избыточные записи удалялись вручную. По согласованию с экспертами в данной области и в соответствии с [14] из 40 признаков были удалены 15 нерелевантных признаков, 8 менее важных признаков, основанных на меньшем коэффициенте усиления [15, 16], и 4 признака, которые имеют только одно значение в записях и более 50 % пропущенных значений. На основе [16] дискретизация была применена для производного атрибута «woman_age», который содержит числовое значение, разбитое на семь групп

(AgeG1[15, 19], AgeG2 [20, 24], AgeG3 [25, 29], AgeG4 [30, 34], AgeG5 [35, 39], AgeG6 [40, 44] и AgeG7 [45, 49]).

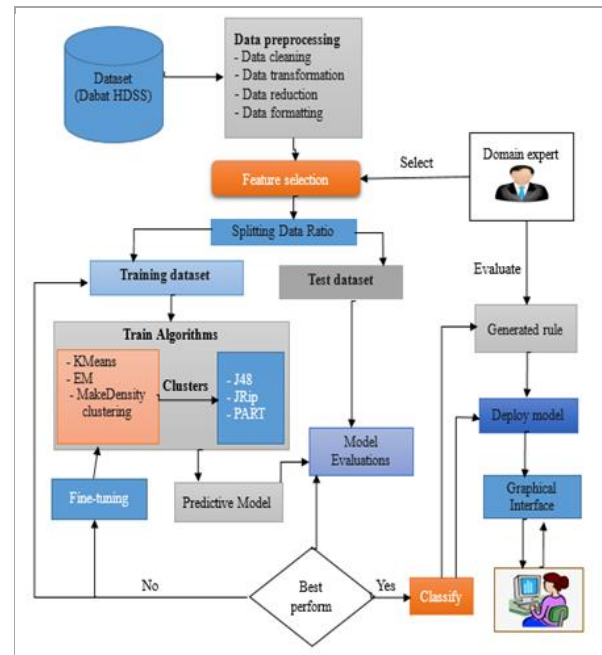


Рис. 1. Архитектура предлагаемой модели

Основываясь на мнении Хана и Камбера [19], мы использовали оценки экспертов, обширный обзор литературы и меньший коэффициент усиления атрибутов, чтобы выбрать несколько полезных атрибутов, игнорируя нерелевантные атрибуты. Таким образом, в табл. I для выбора лучших атрибутов представлены итоговые списки отобранных признаков, использованных в исследовании, в порядке убывания важности на основе коэффициента усиления (Gr) [15] и приблизительного значения информационного усиления (Igr) [16].

ТАБЛИЦА I. ОКОНЧАТЕЛЬНЫЕ СПИСКИ ОТОБРАННЫХ АТТРИБУТОВ

Attr_Name	DT	Gr	Описание	Igr
anc_attend	nm	0.8-1 st	Посещение АНК	0.75-3 rd
hiv_test	nm	0.53-2 nd	ВИЧ-тест	0.53-4 th
anc_place	nm	0.48-3 rd	Место посещения АНК	0.78-1 st
anc_freq	nm	0.35-4 th	Частота АНК	0.77-2 nd
curr_vacc	nm	0.24-5 th	Вакцинация	0.23-5 th
prev_preg_f	nm	0.09-6 th	Частота наступления беременности	0.14-6 th
location	nm	0.08-7 th	Где живут женщины	0.05-9 th
woman_age	nm	0.05-8 th	Возраст женщины	0.115-7 th
educ_categ	nm	0.05-9 th	Категория образования женщин	0.063-8 th
prev_preg_outcome	nm	0.04-10 th	Исход предыдущей беременности	0.023-10 th
m_status	nm	0.03-11 th	Семейное положение матери	0.016-11 th
preg_inten	nm	0.03-12 th	Является ли нынешняя беременность желанной?	0.02-12 th
first_preg	nm	0.02-13 th	Первая беременность	0.006-13 th

IV. ЭКСПЕРИМЕНТЫ И РЕЗУЛЬТАТЫ

В общей сложности 5990 предварительно обработанных записей с 13 выбранными атрибутами используются для формирования моделей кластеризации и классификации в инструменте машинного обучения Waikato Environment for Knowledge Analysis.

A. Кластерное моделирование

Для моделирования кластера были проведены кластеризация K-means, EM и Make-Density с изменением стандартных параметров. K-means выбран для данного исследования из-за легкости реализации, простоты, интерпретируемости, эффективности, эмпирического успеха и популярности. Кластеризация Make-Density выбрана для эксперимента по кластерному моделированию, поскольку позволяет находить кластеры произвольной формы. Это односкановый алгоритм, который должен исследовать исходные данные только один раз, он устойчив к зашумленным данным, в отличие от иерархической кластеризации. EM также выбран потому, что он имеет сильную статистическую основу, устойчив к зашумленным данным, в отличие от иерархической кластеризации. EM также обеспечивает чрезвычайно полезный результат для реальных наборов данных, поскольку мы использовали реальные наборы данных о местном здравоохранении.

Для оценки качества кластера использовались сумма квадратичных ошибок (SSE), количество итераций (i), временная сложность (в секундах) и мнение эксперта в данной области [20].

После обсуждения с экспертами в данной области мы решили сделать k=3 предопределенных кластерных меток (высокий, средний и низкий статус) родового наблюдения, чтобы иметь соответствующие кластеры, основанные на количестве классов. На основе сходства расстояний (количество итераций и Log likelihood) было разработано 6 кластерных моделей (три с использованием k-means, две с использованием Density based и одна с использованием EM). В модели EM внутрикластерное согласие было не очень хорошим, поэтому дальнейший эксперимент не потребовался, что следует из [20].

B. Сравнение кластерных моделей

ТАБЛИЦА II. СРАВНЕНИЕ КЛАСТЕРНЫХ МОДЕЛЕЙ

Cluster Alg	Exp	Distance	S	i	SSE	Sec	Log lh	nFlds
K-means	1	ED	10	3	17094	0.06	-	-
	2	ED	100	5	16963	0.11	-	-
	3	MD	1000	3	17143	0.08	-	-
MDB C	4	ED	10	3	17094	0.9	-8.46	-
	5	ED	100	3	-	0.25	-8.88	-
EM	6	-	100	9	-	0.13	-7.97	10

Key: Exp: Experiment number, Clustering Algorithms, MDB: MakeDensityBasedClusterer, EM: Expectation maximization, ED: EuclideanDistance, MD: ManhattanDistance, Loglh: Log likelihood

Как видно из табл. 2, шестой эксперимент показывает большее количество итераций по сравнению с другими экспериментами, которое больше на четыре по сравнению со вторым экспериментом и на шесть по

сравнению с первым, третьим, четвертым и пятым экспериментами. С другой стороны, в шестом эксперименте логарифм правдоподобия показывает наименьшее значение по сравнению с четвертым и пятым экспериментами. Это указывает на то, что с точки зрения правдоподобия алгоритм EM лучше, чем алгоритм кластеризации MakeDensityBasedClusterer.

Однако в Log likelihood и EM, и MakeDensityBasedClusterer кластеры имеют отрицательные значения, что не свидетельствует о совершенстве результатов и отсутствии необходимости проведения дальнейшего эксперимента.

Таким образом, мы можем сделать вывод, что алгоритм K-means намного лучше, чем EM и MakeDensityBasedClusterer кластеризация по времени построения модели из-за своей простоты. Модель, созданная первым кластером, хорошо сегментирует записи по сравнению с пятью другими кластеризациями, которые имеют меньшее минимальное количество квадратичных ошибок, количество итераций и время построения модели. Кроме того, к экспертам в данной области обращаются, чтобы они высказали свое мнение о том, соответствует ли результат кластеризации бизнесу.

C. Классификационное моделирование

Лучший кластеризованный набор данных используется в качестве входных данных для обучения выбранных алгоритмов классификации. Из доступных атрибутов, используемых в алгоритме кластеризации, 13 используются в качестве независимых переменных, а последние метки классов, полученные в результате кластеризации, – в качестве зависимых атрибутов. Для эксперимента было отобрано 5990 экземпляров, состоящих из 1 699 кластеров 0 (средний), 2 989 кластеров 1 (высокий) и 1 302 кластеров 2 (низкий). 20 % от общего количества записей было выбрано в качестве тестового поднабора данных, а остальные 80 % – в качестве обучающего набора данных, что было принято как эмпирически наилучшее разделение на обучающий и тестовый наборы [14].

В экспериментах использовались древовидные классификаторы J48 и алгоритмы индуктивной классификации JRip и PART. Был проведен 21 эксперимент (восемь моделей с использованием J48, восемь моделей с использованием PART и пять моделей с использованием JRip), в которых рассматривались два случая (усеченный и не усеченный) с двумя схемами (для всех и выбранных атрибутов), три тестовых варианта, которые являются всеми для обучения, 10-кратная кросс-валидация [21] и 80 % процентное разделение. Для оценки результатов также использовались стандартные метрики accuracy, precision, specificity, relative optimal character curve (ROC), mean absolute error, recall and confusion matrix [22] [23]. Кроме того, мы рассмотрели текущие базовые показатели в качестве подходящего теста для оценки статистической значимости предложенной модели

D. Сравнение моделей прогнозирования

ТАБЛИЦА III. СРАВНЕНИЕ МОДЕЛЕЙ ПРОГНОЗА

Модель классификации	Общая точность				Время
	Правильная классификация		Неправильная классификация		
	№	%	№	%	
Unpruned J48 decision tree со всеми атрибутами с использованием 80-ого разбиения	1495	99.87%	2	0.13 %	0.01
Pruned PART rule induction со всеми атрибутами с использованием 80-ого разбиения	1496	99.93%	1	0.07 %	0.07
Pruned JRip rule induction со всеми атрибутами с использованием 80-ого разбиения	1494	99.81%	3	0.19 %	0.27

Как показано в табл. 3, в усеченной индукции правил JRip из всего набора обучающих данных 1497 записей 1494 (99,81 %) записи классифицированы правильно, а три (0,19 %) записи классифицированы неверно. В неусеченном дереве решений J48 из всего набора обучающих данных 1497 записей 1495 (99,87 %) записей классифицированы правильно, а две (0,13 %) записи классифицированы неверно. В усеченной индукции правил PART зарегистрирована точность, из общего набора обучающих данных 1497 записей 1496 (99,93 %) записей классифицированы правильно, и только одна (0,07 %) запись классифицирована неверно. Таким образом, наибольшее количество неверных классификаций зарегистрировано алгоритмом JRip с большим временем вычислений (0,27 сек).

Точность модели PART показывает самую высокую точность предсказания при среднем времени построения модели 0,07 секунды. В табл. 4 показана матрица смешения для выбранной индукционной предсказательной модели PART с усеченными правилами.

ТАБЛИЦА IV. МАТРИЦА СМЕШЕНИЯ

Фактический кластер	Прогнозируемый кластер			Итого	Правильная классификация
	Кластер 0	Кластер 01	Кластер 2		
Кластер 0	420	0	0	420	100%
Кластер 1	0	767	0	767	100%
Кластер 2	0	1	309	310	99.68%
Итого	420	768	309	1497	99.93%

Полученная матрица смешения (табл. 4), кластер 0 (средний статус) показывает результат точности 100 %, это означает, что из 420 записей все 420 записей классифицированы правильно.

В кластере 1 также зарегистрирован результат точности 100 %, что означает, что из общего числа 767 записей все 767 записей правильно классифицированы как кластер 1 (высокий статус). Кластер 2 (низкий статус), в котором зарегистрирован результат точности

99,68 %, содержит в общей сложности 310 записей, из которых 309 записей правильно классифицированы как кластер 2, а 1 запись неправильно классифицирована как кластер 1, в то время как на самом деле является кластером 2. Отсюда можно сделать вывод, что в кластерах 0 и 1 100 % записей правильно классифицированы в назначенный им кластер. Это показывает, что алгоритм классифицировал всех женщин со средним и высоким статусом в соответствующий класс (кластерный индекс). По сравнению с [12] [13], мы использовали двухэтапные подходы машинного обучения, в которых первый алгоритм используется для вменения меток классов набора данных, а второй алгоритм используется для предсказания значений, чтобы получить точный результат [24]. Таким образом, логично заключить, что алгоритм индукции правил J48, PART и JRip более подходит для указанной проблемной области с точностью более 99 %.

Тем не менее, в качестве окончательной рабочей классификационной модели для прогнозирования состояния наблюдения за беременностью женщин была взята модель, разработанная с использованием усеченного PART (наибольшая точность).

E. Примеры интересных правил, сгенерированных алгоритмом PART

Правило 1: ЕСЛИ anc_attend = Да И hiv_test = Да: кластер 1-Высокий статус (2576,0).

Правило 2: ЕСЛИ curr_vacc = Нет И anc_attend = Нет: кластер 2 - Низкий статус (1233,0).

Правило 1 и Правило 2 показывают, что если женщина посещает дородовое наблюдение и обследование на ВИЧ во время беременности, то вероятность того, что женщина попадет в кластер 1 (высокий), отнесенный вышеуказанными предикторами, составляет около 100% с правильным результатом 2576 из 2576 случаев. Если же женщина не делает прививки и вообще не посещает АНК, то вероятность того, что женщина попадет в кластер 2 (низкий), отнесенный к вышеуказанным предикторам, составляет около 100 % с правильным результатом 1233 из 1233.

Правило 3: ЕСЛИ curr_vacc = Да И anc_place = Hpost И anc_freq = Один раз: кластер 0-Средний статус (43,0).

Правило 4: ЕСЛИ curr_vacc = Yes AND anc_place = Hpost AND anc_freq = Twice: кластер 0-Средний статус (40,0). **Правило 5:** ЕСЛИ curr_vacc = Yes AND anc_freq = Thrice AND anc_place = HCenter: кластер 1-Высокий статус (16,0).

Правила 3, 4 и 5, представленные выше, показывают, что если женщина делает прививки и посещает женскую консультацию один или два раза в медицинском учреждении во время текущей беременности, то эта женщина относится к группе со средним статусом наблюдения за беременностью.

Правило 6: ЕСЛИ curr_vacc = Да И woman_age = AgeG3 [25-29] И anc_place = HCenter: кластер1 (71,0). Правило 6 показывает, что если женщина прошла вакцинацию, имеет возрастную группу [25–29] и посещает АНК в центре здоровья, то эта женщина попадает в группу с высоким статусом наблюдения

беременности. Это указывает на то, что демографическая переменная или атрибут возраста женщины также является фактором, влияющим на статус наблюдения беременности.

V. БУДУЩИЕ ИССЛЕДОВАНИЯ

В будущем исследовании предполагается использовать характер занятости женщин, участие мужа, экономический статус и другие параметры инфраструктуры. Необходимо обеспечить интерпретируемость модели и разработать применимую систему поддержки принятия решений, включающую соответствующие средства объяснения в дополнение к прогнозированию статуса, что поможет лучше понять использование женщинами услуг по уходу за беременностью.

VI. ЗАКЛЮЧЕНИЕ

В данном исследовании рассматривается идея использования кластеризации и прогностического моделирования для повышения точности прогностической модели путем сегментирования данных о регистрации женщин с помощью методов машинного обучения. Наиболее важными факторами для прогнозирования беременности оказались атрибуты посещения женской консультации, местонахождение женской консультации, частота посещения женской консультации, иммунизация, тестирование на ВИЧ и возраст женщины. Наиболее важными факторами для прогнозирования статуса наблюдения за беременностью женщин оказались атрибуты посещения женской консультации. Результаты данного исследования позволят принимать обоснованные решения по организации родового наблюдения в странах Азии и Африки.

БЛАГОДАРНОСТЬ

Авторы благодарят Исследовательский центр Дабат при Университете Гондара, Гондар-Эфиопия, за предоставленный набор данных для данного исследования.

СПИСОК ЛИТЕРАТУРЫ

- [1] Всемирная организация здравоохранения (ВОЗ) и ЮНИСЕФ "Обратный отчет до 2015 года: Выживание матерей, новорожденных и детей: построение будущего для женщин и детей: Доклад 2012 года. Женева; 2012 г.
- [2] Всемирная организация здравоохранения, IMPAC Integrated Management of Pregnancy and Childbirth WHO Recommended Interventions for Improving Maternal and Newborn Health. Женева: Всемирная организация здравоохранения, 2010.
- [3] Campbell OMR, Graham WJ, Strategies for reducing maternal mortality: getting on with what works. The Lancet 368: 1284-1299 [PubMed], 2006.
- [4] Всемирная организация здравоохранения (ВОЗ). Antenatal care in развивающихся странах: обещания, достижения и упущенные

возможности: анализ тенденций, уровней и различий, 1990-2001. Публикация в библиотечном каталоге ВОЗ, 2003.

- [5] Гурмеса Тура. "Использование услуг родового ухода и связанные с этим факторы в зоне Метекель, северо-запад Эфиопии" // *Ethiop J Health Sci.*, Vol.19, No. 2, July 2009.
- [6] "Maternal_Health.Internet:"https://www.moh.gov.et/site/initiatives-4-col/Maternal_Health, [Sep 29, 2022].
- [7] Михаэла Опреа, "Об использовании методов обработки данных в системах, основанных на знаниях" // *Экономическая информатика*, апрель 2006 г., том 6, стр. 21-24.
- [8] Шахджахан и др. "Факторы, связанные с использованием услуг родового ухода в сельской местности Бангладеш". *Юго-Восточная Азия // Journal of Public Health*, 2012; 2(2):61-66.
- [9] Bereket Y., Mulat T., Wondimagegn P. "Использование матерями родового ухода и их удовлетворенность услугами родовспоможения в отдельных государственных медицинских учреждениях зоны Wolaita, Южная Эфиопия // *IJSTR*, 2013; 2(2): 74-85.
- [10] Кассу Мехари. "Детерминантные факторы, влияющие на использование услуг по охране материнского здоровья в сельских районах Эфиопии". Магистерская диссертация, Университет Аддис-Абебы, Эфиопия, июнь 2012 г.
- [11] Теводрос А., Джемаль Х., Дередже Х. Использование услуг родового ухода среди подростков в Эфиопии: Перекрестное секционное исследование // *Ethiopia Journal of Health Development*, 24(3):221-225, 2010.
- [12] Тарику Дебела. "Разработка прогностической модели фертильных предпочтений женщин репродуктивного возраста с использованием технологии интеллектуального анализа данных". Магистерская диссертация, Университет Аддис-Абебы, Эфиопия, 2013.
- [13] Теклехайманот Асефа. "Прогнозирование наблюдения за женщинами во время беременности в Эфиопии с помощью технологии Data Mining". M.Sc. thesis, Addis Ababa University, Ethiopia, October, 2015.
- [14] Дж. Хан и М. Камбер, "Data Mining: Concepts and Techniques," Second Edition, Morgan Kauffman Publishers, San Francisco.
- [15] Cios Krzysztof J, Pedrycz Witold, Swiniarski Roman W, Kurgan Lukasz A. Data Mining: A Knowledge Discovery Approach. Нью-Йорк, США: Springer Science Business Media LLC; 2007.
- [16] Куинлан Дж. R, C4.5: Программы для машинного обучения. Los Altos: Morgan Kaufmann, 1993.
- [17] [Online]. Available: <http://www.indepthishare.org/index.php/catalog/122>. [Accessed 18 January 2022].
- [18] [Online]. Available: <http://drc.uog.edu.et> [Accessed 10 October 2022].
- [19] Хан Дж. и Камбер М. Добыча данных: Concepts and Techniques. New York. США: Morgan Kaufmann; 2001 г.
- [20] С. Теодоридис и К. Кутробас. Распознавание образов. Academic Press, 1999.
- [21] Тесфахун Хайлемариам. "Применение Data Mining для прогнозирования смертности среди взрослых". M.Sc. Thesis, Университет Аддис-Абебы, июнь, 2012.
- [22] Использование матрицы путаницы для улучшения ансамблевых классификаторов, Инженеры по электротехнике и электронике в Израиле (IEEE), 2010 IEEE 26-й съезд Дата проведения конференции: 17-20 ок., Page(s):000555-000559, 2010.
- [23] М. Кантарджич, Концепции, модели, методы и алгоритмы добычи данных. John wiley and Sons publication Inc: США, 2003.
- [24] A Beginner. Guide to Two-Phase Learning: <https://analyticsindiamag.com/a-beginnersguide-to-two-phase-learning/>, [Oct 05, 2023].