Федеральное государственное образовательное бюджетное учреждение высшего образования

«Финансовый университет при Правительстве Российской Федерации»
(Финансовый университет)

Бузулукский филиал Финуниверситета

УТВЕРЖДАЮ
Директор Бузулукского
филиала Финуниверситета
_____ Н. Н. Щавелева
«20» октября 2026 г.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для государственной итоговой аттестации
по специальности 09.02.13 Интеграция решений с применением технологий искусственного интеллекта

Бузулук -2026 г.

Фонд оценочных средств для государственной итоговой аттестации разработан на основе федерального государственного образовательного стандарта среднего профессионального образования по специальности 09.02.13 Интеграция решений с применением технологий искусственного интеллекта

Составители:

Петрянкина Елена Сергеевна, преподаватель, высшая квалификационная категория

Фонд оценочных средств для государственной итоговой аттестации рассмотрен и рекомендован к утверждению на заседании предметно-цикловой комиссии экономических, учетных и кредитных дисциплин

Протокол №3 от «29» октября 2026 г.
Председатель предметно-цикловой комиссии  Е.С. Петрянкина

Согласовано:

ООО «КЭтрион», директор Этманова Екатерина Александровна

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ГОСУДАРСТВЕННОЙ ИТОГОВОЙ АТТЕСТАЦИИ

Фонд оценочных средств для государственной итоговой аттестации (далее – ГИА) разработан на основе:

– Приказ Министерства просвещения Российской Федерации от 24.12.2024 № 1025 «Об утверждении федерального государственного образовательного стандарта среднего профессионального образования по специальности 09.02.13 «Интеграция решений с применением технологий искусственного интеллекта» (зарегистрирован Министерством юстиции Российской Федерации 27 января 2025 г., регистрационный №81046);

– Приказ Минпросвещения России от 08 ноября 2021 г. №800 «Об утверждении Порядка проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования» (зарегистрирован Министерством юстиции Российской Федерации 07 декабря 2021 г., регистрационный № 66211);

– Приказ Минобрнауки России и Минпросвещения России от 05 августа 2020 г. № 885 / 390 «О практической подготовке обучающихся» (зарегистрирован Министерством юстиции Российской Федерации 11 сентября 2020 г., регистрационный № 59778);

– Приказ Министерства труда и социальной защиты Российской Федерации от 05 сентября 2017 года № 658н «Об утверждении профессионального стандарта «Специалист по интеграции прикладных решений» (зарегистрирован Министерством юстиции Российской Федерации 22 сентября 2017 года, рег.№ 48309).

В рамках специальности СПО предусмотрено освоение квалификации «специалист по работе с искусственным интеллектом».

Выпускник, освоивший ППССЗ, должен быть готов к выполнению основных видов деятельности согласно получаемой квалификации «специалист по работе с искусственным интеллектом».

В соответствии с основными видами деятельности, предусмотренными ФГОС СПО, в ППССЗ входят профессиональные модули:

ПМ.01 Разработка кода для искусственного интеллекта

ПМ.02 Администрирование баз данных

ПМ.03 Обучение готовых моделей искусственного интеллекта

1.1. Область применения оценочных средств государственной итоговой аттестации

Оценочные задания по квалификации «специалист по работе с искусственным интеллектом» разработаны с учетом соответствия основных видов деятельности и профессиональных компетенций, установленных ФГОС СПО. Результаты освоения ППССЗ определяются приобретаемыми выпускником компетенциями, т.е. его способностью применять знания, умения и личные качества в соответствии с задачами профессиональной деятельности. В результате освоения ППССЗ выпускник должен обладать следующими общими и профессиональными компетенциями:

ОК 01 Выбирать способы решения задач профессиональной деятельности применительно к различным контекстам

ОК 02 Использовать современные средства поиска, анализа и интерпретации информации и информационные технологии для выполнения задач профессиональной деятельности

ОК 03 Планировать и реализовывать собственное профессиональное и личностное развитие, предпринимательскую деятельность в профессиональной сфере, использовать знания по правовой и финансовой грамотности в различных жизненных ситуациях

ОК 04 Эффективно взаимодействовать и работать в коллективе и команде

ОК 05 Осуществлять устную и письменную коммуникацию на государственном языке Российской Федерации с учетом особенностей социального и культурного контекста

ОК 06 Проявлять гражданско-патриотическую позицию, демонстрировать осознанное поведение на основе традиционных российских духовно-нравственных ценностей, в том числе с учетом гармонизации межнациональных и межрелигиозных отношений, применять стандарты антикоррупционного поведения;

ОК 07 Содействовать сохранению окружающей среды, ресурсосбережению, применять знания об изменении климата, принципы бережливого производства, эффективно действовать в чрезвычайных ситуациях

ОК 08 Использовать средства физической культуры для сохранения и укрепления здоровья в процессе профессиональной деятельности и поддержания необходимого уровня физической подготовленности

ОК 09 Пользоваться профессиональной документацией на государственном и иностранном языках

ВД.1 Разработка кода для искусственного интеллекта

ПК 1.1. Формировать алгоритмы разработки программных модулей в соответствии с техническим заданием.

ПК 1.2. Разрабатывать программные модули в соответствии с техническим заданием.

ПК 1.3. Оформлять программный код в соответствии с техническим заданием.

ПК 1.4. Использовать систему контроля версий программного кода с учетом обеспечения возможности организации групповой разработки.

ПК 1.5. Выполнять отладку программных модулей с использованием специализированных программных средств.

ПК 1.6. Выполнять тестирование программного кода.

ПК 1.7. Составлять тестовые сценарии.

ВД.2 Администрирование баз данных

ПК 2.1 Выявлять проблемы, возникающие в процессе эксплуатации баз данных.

ПК 2.2 Осуществлять процедуры администрирования баз данных.

ПК 2.3 Проводить аудит систем безопасности баз данных с использованием регламентов по защите информации.

ПК 2.4 Формировать требования хранилищ банка данных для обучения.

ПК 2.5 Подготавливать данные для базы знаний

ВД 3 Обучение готовых моделей искусственного интеллекта

ПК 3.1 Осуществлять выбор готовых моделей искусственного интеллекта.

ПК 3.2 Формировать сценарии обучения готовых моделей искусственного интеллекта.

ПК 3.3 Проводить обучение и последующую калибровку готовых моделей искусственного интеллекта.

ПК 3.4 Контролировать результат обучения.

ПК 3.5 Оформлять результат проведения процедуры обучения.

ПК 3.6 Формировать запросы для работы с искусственным интеллектом с целью визуализации данных.

1.2. Цели и задачи государственной итоговой аттестации

Целью государственной итоговой аттестации является установление соответствия уровня освоения компетенций, обеспечивающих соответствующую квалификацию и уровень образования обучающихся, ФГОС СПО по специальности 09.02.13 «Интеграция решений с применением технологий искусственного интеллекта». ГИА призвана способствовать систематизации и закреплению знаний и умений обучающегося по специальности при решении конкретных профессиональных задач, определить уровень подготовки выпускника к самостоятельной работе.

2. СТРУКТУРА ПРОЦЕДУРЫ ГИА И ПОРЯДОК ЕЕ ПРОВЕДЕНИЯ

2.1. Структура задания для процедуры ГИА

В соответствии с ФГОС СПО ГИА по ППССЗ 09.02.13 Интеграция решений с

применением технологий искусственного интеллекта проводится в форме демонстрационного экзамена. Структура задания для процедуры ГИА включает выполнение каждым обучающимся заданий демонстрационного экзамена базового уровня.

ДЭ направлен на определение уровня освоения выпускником материала, предусмотренного образовательной программой, и степени сформированности профессиональных умений и навыков путем проведения независимой экспертной оценки выполненных выпускником практических заданий в условиях реальных или смоделированных производственных процессов.

ДЭ в рамках ГИА проводится с использованием КИМ, представляющих собой комплексы заданий стандартизированной формы, и критериев их оценивания.

КИМ, критерии оценивания доводятся оператором до члена государственной экзаменационной комиссии, присутствующего на демонстрационном экзамене, за один рабочий день до дня проведения демонстрационного экзамена.

2.2. Порядок проведения процедуры государственной итоговой аттестации

В соответствии с пунктом 12 Порядка проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования, утверждённого приказом Министерства просвещения Российской Федерации от 8 ноября 2021 г. № 800 (в редакции приказа Министерства просвещения Российской Федерации от 22 мая 2026 г. № 351) (далее – Порядок), государственная итоговая аттестация (далее – ГИА) в форме демонстрационного экзамена (далее – ДЭ) проводится единой государственной экзаменационной комиссией (далее – ГЭК), создаваемой исполнительным органом субъекта Российской Федерации, осуществляющим государственное управление в сфере образования, – министерством образования Оренбургской области (далее – министерство).

В соответствии с пунктом 14 Порядка состав ГЭК утверждается распорядительным актом министерства, не позднее 15 ноября года, предшествующего году проведения ГИА. В состав ГЭК входят председатель, заместители председателя, члены и ответственный секретарь. Председателем ГЭК является руководитель или заместитель руководителя министерства.

В соответствии с пунктом 74 Порядка при проведении ГИА в форме ДЭ состав апелляционной комиссии утверждается одновременно с утверждением состава ГЭК.

ГЭК формируется из числа педагогических и иных работников образовательных организаций, лиц, приглашённых из сторонних организаций, в том числе (п. 12 Порядка):

– педагогических и иных работников образовательной организации;

– представителей работодателей, направление деятельности которых соответствует области профессиональной деятельности, к которой готовятся выпускники.

Формирование состава ГЭК осуществляется исходя из необходимости обеспечения присутствия не менее одного её члена на каждом ДЭ в соответствии с графиком проведения ДЭ субъекта Российской Федерации (п. 14 Порядка).

Функции члена ГЭК, присутствующего на демонстрационном экзамене:

– получение от оператора КИМ, критериев оценивания (за 1 рабочий день до дня проведения демонстрационного экзамена) (пункт 26 Порядка);

– ознакомление с КИМ, критериями оценивания администратора ЦПДЭ, экспертов, технического эксперта (не позднее, чем за 1 день до проведения демонстрационного экзамена) (пункт 37 Порядка);

– наблюдение за ходом проведения демонстрационного экзамена (пункт 37 Порядка);

– присутствие на выставлении баллов в протоколе проведения демонстрационного экзамена (пункт 62 Порядка);

– выставление в день проведения закрытого заседания ГЭК оценок по итогам государственной итоговой аттестации и передача оригинала протокола проведения ДЭ на хранение в образовательную организацию (пункты 60, 65, 66 Порядка).

Апелляционная комиссия состоит из председателя апелляционной комиссии, не менее пяти членов апелляционной комиссии и секретаря апелляционной комиссии из числа педагогических работников образовательной организации, не входящих в данном учебном году в состав ГЭК (п. 74 Порядка).

Председателем апелляционной комиссии может быть назначено лицо из числа руководителей или заместителей руководителей организаций, осуществляющих образовательную деятельность, соответствующую области профессиональной деятельности, к которой готовятся выпускники, представителей работодателей или их объединений, включая экспертов, при условии, что направление деятельности данных представителей соответствует области профессиональной деятельности, к которой готовятся выпускники, и при условии, что такое лицо не входит в состав ГЭК (п. 74 Порядка).

В соответствии с пунктом 16 Порядка проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования, утверждённого приказом Министерства просвещения Российской Федерации от 8 ноября 2021 г. № 800 (в редакции приказа Министерства просвещения Российской Федерации от 22 мая 2026 г. № 351) (далее – Порядок), для проведения демонстрационного экзамена на территории субъекта Российской Федерации решением государственной экзаменационной комиссии (далее – ГЭК) на основании предложений образовательных организаций в срок не позднее 15 декабря года, предшествующего календарному году проведения демонстрационного экзамена, утверждается список экспертов по укрупнённым группам профессий и специальностей.

В соответствии с пунктом 16 Порядка эксперты утверждаются из числа педагогических работников, представителей работодателей и их объединений, обладающих профессиональными знаниями, навыками и опытом в сфере, соответствующей профилю УГПС, по которой проводится демонстрационный экзамен.

В соответствии с пунктом 17 Порядка:

– экспертная группа создается по каждой профессии и специальности;

– в состав экспертной группы не могут входить эксперты, имеющие трудовые или иные договорные отношения с образовательной организацией, в которой проводится демонстрационный экзамен;

– эксперты должны соответствовать профилю профессии, специальности среднего профессионального образования или виду деятельности, по которому проводится демонстрационный экзамен.

Функции членов экспертной группы:

– исполнение обязанностей по оценке выполнения КИМ, распределенных администратором ЦПДЭ (пункт 31 Порядка);

– осуществление оценки выполнения КИМ (пункт 38 Порядка);

– фиксация результатов выполнения выпускниками КИМ в соответствии с критериями оценивания (пункт 57 Порядка);

– подписание протокола проведения демонстрационного экзамена (пункт 62 Порядка).

В соответствии с пунктом 24 Порядка проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования, утверждённого приказом Министерства просвещения Российской Федерации от 8 ноября 2021 г. № 800 (в редакции приказа Министерства просвещения России от 22 мая 2026 г. № 351), полномочия по формированию графика проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования в форме демонстрационного экзамена в субъекте Российской Федерации определены исполнительному органу субъекта Российской Федерации, осуществляющему государственное управление в сфере образования, – министерству образования Оренбургской области.

Сроки проведения каждой формы ГИА регламентируются филиалом в соответствии с учебным календарным графиком учебного процесса на текущий учебный год.

Для выполнения демонстрационного экзамена необходимо соблюдение следующих

условий:

- демонстрационный экзамен проводится на площадке, аккредитованной в качестве центра проведения демонстрационного экзамена;

- площадка оснащается в соответствии с установленными требованиями по выбранного кода.

Образовательная организация обеспечивает необходимые технические условия для обеспечения КИМ, критериями оценивания во время демонстрационного экзамена выпускников, членов государственной экзаменационной комиссии, администратора центра проведения экзамена, экспертов.

Обучающиеся проходят ДЭ в ЦПДЭ в составе экзаменационных групп.

Образовательная организация знакомит с планом проведения демонстрационного экзамена выпускников, сдающих демонстрационный экзамен и лиц, обеспечивающих проведение демонстрационного экзамена в срок не позднее чем за пять рабочих дней до даты проведения экзамена.

Количество, общая площадь и состояние помещений, предоставляемых для проведения ДЭ, должны обеспечивать проведение ДЭ в соответствии со спецификацией КИМ.

Не позднее чем за один рабочий день до даты проведения ДЭ администратором центра проведения экзамена проводится проверка готовности ЦПДЭ в присутствии членов экспертной группы, выпускников, а также технического эксперта, назначаемого организацией, на территории которой расположен центр проведения экзамена, ответственного за соблюдение установленных норм и правил охраны труда и техники безопасности.

Администратором центра проведения экзамена осуществляется осмотр центра проведения экзамена, распределение обязанностей между членами экспертной группы по оценке выполнения КИМ, а также распределение рабочих мест между выпускниками с использованием способа случайной выборки. Результаты распределения обязанностей между членами экспертной группы и распределения рабочих мест между выпускниками фиксируются администратором центра проведения экзамена в протоколах.

Допуск выпускников в центр проведения экзамена осуществляется администратором центра проведения экзамена на основании документов, удостоверяющих личность.

Образовательная организация обязана не позднее чем за один рабочий день до дня проведения ДЭ уведомить администратора центра проведения экзамена об участии в проведении ДЭ тьютора (ассистента).

Для выполнения заданий КИМ не предусматривается наличие (присутствие) добровольцев (волонтеров).

Продолжительность ДЭ базового уровня составляет 3 часа 30 минут.

Единое базовое ядро содержания КИМ сформировано на основе видов деятельности в соответствии с ФГОС СПО и является общей содержательной основой заданий ДЭ вне зависимости от вида аттестации и уровня ДЭ и содержится в спецификации КИМ 09.02.13-1-2027 Специалист по работе с искусственным интеллектом (https://bom.firpo.ru/Public/6739)

2.3 Требования к оцениванию.

Распределение значений максимальных баллов зависит от вида аттестации, уровня ДЭ, составной части КИМ. Для государственной итоговой аттестации в форме демонстрационного экзамена базового уровня инвариантной частью КИМ максимальный балл составляет 75 из 75. Распределение баллов по критериям оценивания для ДЭ БУ в рамках ГИА представлено в спецификации КИМ 09.02.13-1-2027 Специалист по работе с искусственным интеллектом (https://bom.firpo.ru/Public/6739)

Количественный состав экспертной группы определяется образовательной организацией, исходя из числа сдающих одновременно ДЭ обучающихся. Один эксперт должен иметь возможность оценить результаты выполнения обучающимися задания в полной мере согласно критериям оценивания.

Перевод результатов демонстрационного экзамена в пятибалльную оценку разработан на основе методических рекомендаций в рамках совершенствования практики применения пункта 60 Порядка проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования, утвержденного приказом Минпросвещения России от 8 ноября 2021 г. № 800 «Об утверждении Порядка проведения государственной итоговой аттестации по образовательным программам среднего профессионального образования».

Согласно пункту 60 Порядка проведения ГИА СПО результаты проведения ГИА оцениваются с проставлением одной из отметок: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно» – и объявляются в тот же день после оформления протоколов заседаний государственных экзаменационных комиссий.

В целях обеспечения сопоставимости качественной и общей успеваемости обучающихся по результатам демонстрационного экзамена, а также в связи с необходимостью обеспечения преемственности и сравнения образовательных результатов в субъектах Российской Федерации образовательным организациям при подготовке локальных нормативных актов, регламентирующих перевод результатов демонстрационных экзаменов в баллах в отметки по пятибалльной системе оценивания, **рекомендуется использовать для всех видов аттестации и уровней демонстрационных экзаменов шкалу перевода «0–50–65–90–100».**

Данная шкала предусматривает следующее соотношение между баллами демонстрационного экзамена, полученными за выполнение практических заданий, и отметками по пятибалльной системе оценивания (см. таблицу 1).

Таблица 1. Рекомендованная шкала перевода баллов демонстрационного экзамена в отметки по пятибалльной системе оценивания

Оценка	**Неудовлетворительно «2»**	**Удовлетворительно «3»**	**Хорошо «4»**	**Отлично «5»**
Отношение полученного количества баллов к максимально возможному (в процентах)	**0,00-49,99%**	**50,00-64,99%**	**65,00-89,99%**	**90,00-100%**
Количество баллов, полученных при сдаче ДЭ в рамках промежуточной аттестации (максимальный балл 25)	0-12,4	12,5-16,2	16,3-22,4	22,5-25
Количество баллов, полученных при сдаче ДЭ базового уровня (максимальный балл 50)	0-24,9	25-32,4	32,5-44,9	45-50
Количество баллов, полученных при сдаче ДЭ профильного уровня (максимальный балл 75)	0-37,4	37,5-48,6	48,7-67,4	67,5-75
Количество баллов, полученных при сдаче ДЭ ПУ - совокупность инвариантной и вариативной частей (25 баллов) (максимальный балл 100)	0-49,9	50-64,9	65-89,9	90-100

При использовании данной шкалы только более половины выполненных заданий гарантирует обучающему успешную сдачу демонстрационного экзамена. Например, на итоговой аттестации в форме демонстрационного экзамена профильного уровня без вариативной части (максимально возможное количество баллов – 80) выпускник считается сдавшим экзамен при количестве 40 и более набранных баллов.

Применение данной шкалы повышает требования к качеству результатов реализации образовательных программ СПО и, тем самым, отвечает государственной политике в системе профессионального образования.

2.4 Образец задания для ГИА ДЭ БУ

Задание 1. Разработка алгоритма и обучение модели искусственного интеллекта В рамках задания Вам необходимо разработать систему анализа тональности текстовых отзывов, способную классифицировать сообщение пользователя как positive (положительное) или negative (отрицательное). Исходные данные представлены в приложении и содержат сами отзывы и соответствующие им метки классов. В приложении приведён список стоп-слов русского языка, которые необходимо исключить при обработке текста. На первом этапе требуется выполнить проектирование алгоритма предварительной обработки текста. Вам необходимо продумать последовательность преобразований, которые позволяют превратить «сырой» текст отзыва в набор информативных признаков, пригодных для использования в модели машинного обучения. Как правило, данный процесс включает следующие шаги:

• приведение текста к единому регистру;

• удаление пунктуации и специальных символов;

• разбиение текста на слова (токенизация);

• удаление стоп-слов из списка, представленного в приложении;

• при необходимости — лемматизацию или стемминг (если это улучшает качество модели);

• оценка полученных результатов и их визуализация

При описании алгоритма важно не ограничиваться перечислением операций. Для каждого этапа необходимо кратко пояснить:

• какую задачу он решает;

• почему он необходим в контексте данной задачи;

• как он влияет на качество итоговой модели.

Результат описания необходимо оформить либо в комментариях к коду, либо в отдельном файле Алгоритм.docx. После определения алгоритма следует перейти к его программной реализации. Необходимо разработать отдельный модуль на языке Python (или любом другом языке, подходящем для задач машинного обучения), который должен быть сохранён под именем classifier. Внутри модуля должны быть реализованы как минимум две основные функции.

1. Функция загрузки и предобработки данных Данная функция должна обеспечивать:

• загрузку данных из CSV-файла;

• применение описанного алгоритма предобработки ко всем текстам;

• формирование выходных данных в виде:

• очищенных текстов;

• соответствующих меток классов.

2. Функция построения и обучения модели Данная функция должна реализовывать полный цикл обучения модели и включать следующие этапы:

• разделение данных на обучающую и тестовую выборки в соотношении 80/20;

• обучение TF-IDF векторизатора только на обучающей выборке;

• преобразование обучающей и тестовой выборок в числовое представление;

• обучение модели логистической регрессии из библиотеки scikit-learn (или аналога) с параметром C = 1.0;

• получение предсказаний на тестовой выборке;

• оценку качества модели.

Оценка качества модели должна включать следующие метрики:

• accuracy;

precision;

• recall;

• F1-score.

Допускается использование как метрик по каждому классу, так и усреднённых значений (например, macro- или weighted-average). При этом строго обязательно, чтобы тестовая выборка не использовалась на этапе обучения ни в каком виде — она должна применяться только после обучения модели и векторизатора. Необходимо провести тестирование, позволяющее проверить точность модели. При реализации кода необходимо соблюдать принципы чистого кода. В частности:

• использовать осмысленные имена переменных и функций;

• избегать дублирования логики;

• разделять ответственность между функциями;

• добавлять docstring к ключевым компонентам;

• использовать комментарии только в тех местах, где логика не является очевидной из кода.

В рамках выполнения задания демонстрационного экзамена Вам предстоит разработать систему анализа тональности текстовых отзывов, способную определять, является ли отзыв positive (положительным) или negative (отрицательным). Исходные данные находятся в Приложении 1 и содержат сами отзывы и соответствующие им метки. В Приложении 2 представлен список стоп-слов русского языка, которые необходимо исключить при обработке текста.

Первый этап — проектирование алгоритма предварительной обработки текста. На этом этапе Вам необходимо определить, какие преобразования должен пройти исходный текст, чтобы быть преобразованным в набор информативных признаков, пригодных для машинного обучения. Как правило, процесс включает следующие шаги:

• приведение текста к единому регистру;

• удаление пунктуации и специальных символов;

• разбиение текста на слова (токенизация);

• удаление стоп-слов из списка, представленного в Приложении 2;

• при необходимости — лемматизация или стемминг (если Вы считаете это целесообразным для повышения качества модели);

• оценка полученных результатов и их визуализация Важно не просто перечислить этапы обработки, а обосновать каждый из них, объяснив:

• какую проблему он решает;

• как он влияет на качество признаков;

• почему он важен для задачи классификации.

Описание алгоритма необходимо разместить либо в комментариях к коду, либо в отдельном файле Алгоритм.docx. После формализации алгоритма следует перейти к его программной реализации. Необходимо разработать отдельный программный модуль на языке Python (или любом ином языке программирования, подходящем для задач машинного обучения), который должен быть сохранён под именем classifier. Внутри данного модуля должны быть реализованы как минимум две основные функции.

Первая функция — подготовка данных Она должна обеспечивать полный цикл работы с исходными данными и включать следующие этапы:

• загрузку датасета из CSV-файла;

• применение разработанного алгоритма предобработки ко всем текстовым данным;

• разделение выборки на обучающую и тестовую части в соотношении 80/20;

использование фиксированного параметра случайности (random_state) для обеспечения воспроизводимости результатов.

Вторая функция — обучение модели

Она отвечает за построение и обучение модели машинного обучения и должна включать следующие этапы:

• преобразование текстов в числовое представление с использованием TFIDF векторизации (или её аналога);

• обучение векторизатора только на обучающей выборке;

• преобразование обучающей и тестовой выборок с использованием обученного векторизатора;

• обучение модели логистической регрессии из библиотеки scikit-learn (или аналога).

После обучения модели, необходимо провести тестирование, позволяющее проверить точность модели и выявить базовые метрики. При реализации важно строго соблюдать корректное разделение этапов обучения и тестирования, исключая любое использование тестовой выборки на этапе обучения модели или векторизатора. Требуется сохранить обученную модель и обученный векторизатор на диск, чтобы в дальнейшем их можно было использовать повторно без необходимости переобучения.

Необходимые приложения:

Прил_1_ОЗ_КИМ_09.02.13-1-2027.csv

Прил_2_ОЗ_КИМ_09.02.13-1-2027.txt

Задание 2. Реализация хранения информации в базе данных

В рамках доработки предыдущего задания необходимо расширить систему анализа текстовых отзывов, добавив хранение данных в базе данных и обновлённый пайплайн обучения модели. На первом этапе требуется спроектировать простую структуру базы данных для хранения отзывов. В рамках этой структуры должна быть предусмотрена таблица, в которой сохраняются как исходные тексты отзывов, так и их предобработанные версии. После загрузки данных из файла CSV необходимо выполнить следующие шаги:

• сохранить исходные тексты отзывов в базу данных;

• применить ранее разработанный алгоритм предобработки текста;

• сохранить результат предобработки в отдельное поле той же таблицы.

Таким образом, база данных становится основным источником данных для дальнейшей работы, а прямое использование CSV-файла исключается.

Далее необходимо реализовать процесс извлечения данных из базы и подготовки их к обучению модели. Этот этап включает последовательность действий:

• извлечение данных из базы данных;

• формирование обучающей и тестовой выборок в соотношении 80/20;

• обучение TF-IDF векторизатора только на тренировочной части данных;

• преобразование обучающей и тестовой выборок в числовое представление;

• обучение модели логистической регрессии на полученных признаках.

Дополнительно требуется повысить качество предсказаний модели за счёт калибровки вероятностей. Для этого используется метод CalibratedClassifierCV (или аналог), который позволяет скорректировать вероятностные оценки модели. В рамках этого этапа необходимо:

• выполнить калибровку обученной модели;

• получить более надёжные вероятностные предсказания;

• сохранить итоговую модель и TF-IDF векторизатор на диск для последующего использования. Также необходимо разработать тестовые сценарии, проверяющие корректность работы ключевых компонентов системы.

Тестирование должно включать:

• проверку загрузки данных из CSV-файла;

• проверку загрузки данных из базы данных;

• проверку корректности работы алгоритма предобработки текста;

• проверку процесса обучения модели и получения предсказаний. После обучения модели выполняется оценка её качества.

Для этого используются стандартные метрики классификации:

• accuracy;

• precision;

• recall;

• F1-score.

Дополнительно требуется построить матрицу ошибок (confusion matrix) для визуального анализа результатов работы модели. При реализации решения необходимо соблюдать принципы чистого кода. В частности:

• использовать осмысленные имена функций и переменных;

• разделять логику на независимые компоненты;

• добавлять docstring к ключевым функциям;

• избегать избыточных комментариев, оставляя пояснения только для неочевидной логики. Отдельной частью работы является обоснование принятых решений.

В файле Обоснование_выбора.docx необходимо описать:

• выбор модели (логистическая регрессия);

• обоснование параметров регуляризации;

• необходимость калибровки вероятностей;

• выбор TF-IDF как метода векторизации текста;

• основная информация по работе с базой данных и способы устранения ошибок.

Все созданные файлы должны быть сохранены в системе контроля версий в соответствующей ветке проекта.

Необходимые приложения: отсутствуют.