

МЕЖДУНАРОДНЫЙ ЧЕМПИОНАТ
ВЫСОКОТЕХНОЛОГИЧНЫХ ПРОФЕССИЙ
«ХАЙТЕК: НАВЫКИ БУДУЩЕГО» 2025

КОНКУРСНОЕ ЗАДАНИЕ

**«МАШИННОЕ ОБУЧЕНИЕ
И БОЛЬШИЕ ДАННЫЕ»**

Конкурсное задание включает в себя следующие разделы:

1.	ФОРМА УЧАСТИЯ В КОНКУРСЕ	3
2.	ОБЩЕЕ ВРЕМЯ НА ВЫПОЛНЕНИЕ ЗАДАНИЯ:.....	3
3.	ЗАДАНИЕ ДЛЯ КОНКУРСА	3
4.	МОДУЛИ ЗАДАНИЯ И НЕОБХОДИМОЕ ВРЕМЯ	4
5.	КРИТЕРИИ ОЦЕНКИ	7

1. ФОРМА УЧАСТИЯ В КОНКУРСЕ

Индивидуальный формат участия;

Очная форма проведения с возможность применения дистанционных технологий

2. ОБЩЕЕ ВРЕМЯ НА ВЫПОЛНЕНИЕ ЗАДАНИЯ: 15 ч.

3. ЗАДАНИЕ ДЛЯ КОНКУРСА

В рамках конкурсного задания конкурсантам предстоит работать с большими наборами разнородных данных, которые могут включать табличные структуры, текстовую информацию, изображения и другие форматы. Конкурсантам необходимо будет выполнить комплексную предварительную обработку данных, включая их очистку, нормализацию и преобразование к виду, пригодному для машинного обучения, с учетом специфики каждого типа данных.

На следующем этапе участникам потребуется проанализировать подготовленные данные, определить наиболее значимые признаки и выбрать модели и методы машинного обучения в зависимости от характера решаемой задачи и природы данных и реализовать их. Необходимо провести всестороннюю оценку качества построенных моделей, используя соответствующие метрики, а также выполнить оценку качества на независимой тестовой выборке.

Заключительная часть задания включает разработку программного интерфейса для практического использования созданного решения: участникам предстоит реализовать API для интеграции модели, а также создать удобный графический интерфейс для конечных пользователей. Дополнительно требуется подготовить документацию, описывающую все этапы работы, принятые решения и инструкции по развертыванию и использованию системы.

Конкурсное задание является закрытым и доступно для ознакомления экспертам в день С-1, участникам – в дни С1 - С3 за 15 минут до начала модуля.

4. МОДУЛИ ЗАДАНИЯ И НЕОБХОДИМОЕ ВРЕМЯ

Наименование модуля		Соревновательный день (C1, C2, C3)	Время на задание
A	Парсинг и предобработка данных	C1	3 часа
B	Разведочный анализ данных.	C1	3 часа
C	Построение, обучение и оптимизация модели	C2	3 часа
D	Разработка программного продукта. Разработка средств интеграции и поддержки готового решения.	C2	3 часа
E	Итоговая оценка модели(ей). Разработка пакета документации. Презентация результатов работы.	C3	3 часа

Модуль А: Парсинг и предобработка данных.

Данные представлены в одном из форматов: txt, pdf, doc, docx, xls, xlsx, xml, csv.

Необходимо преобразовать исходный набор данных к виду, пригодному для загрузки.

Примерные этапы задания в модуле А:

- 1) Предобработка данных и выделение значимых атрибутов;
- 2) Разбиение сложных атрибутов;
- 3) Очистка данных от служебной информации.

Модуль В: Разведочный анализ данных.

Необходимо провести предобработку данных, например, разбить сложные атрибуты, очистить от технической информации, удалить спецсимволы. В обработанном наборе данных выделяются значимые атрибуты для дальнейшего анализа и построения модели машинного обучения.

Примерные этапы задания в модуле В:

- 1) Дополнение недостающими данными;
- 2) Формирование словарей (мешков слов);
- 3) Предварительный анализ данных.

Модуль С: Построение, обучение и оптимизация модели.

В этом модуле продолжается работа с данными, подготовленными в предыдущем модуле. На основе предобработанных данных предстоит построить модель машинного обучения. Как правило это модель кластеризации, классификации или регрессии. Необходимо определить точность работы модели и предпринять какие-либо действия для повышения точности.

Примерные этапы задания в модуле С:

- 1) Визуализация зависимостей данных;
- 2) Выбор модели классификации;
- 3) Обучение модели классификации;
- 4) Feature Engineering.

Модуль D: Разработка программного продукта. Разработка средств интеграции и поддержки готового решения.

В этом модуле продолжается работа с данными, подготовленными в предыдущем модуле. Необходимо разработать программный продукт (ПП). Это может быть графический интерфейс, веб-приложение или чат-бот – в зависимости от требований задачи.

Цель – взаимодействие с пользователем на естественном языке, реализация построенной модели для решения прикладных задач пользователя – прогнозирование, классификация, поддержка принятия решений.

Для реализованного программного продукта предстоит разработать интерфейс сбора данных и реализовать возможность повышения точности работы модели. Для этого разрабатывается ряд методов API, способных принимать/возвращать от какой-либо системы данные в установленном формате.

Примерные этапы задания в модуле D:

- 1) Разработка ПП;
- 2) Настройки ПП;

3) Разработка API, интеграция с ПП.

Модуль Е: Итоговая оценка модели(ей). Разработка пакета документации.

Презентация результатов работы.

В этом модуле продолжается работа с программным продуктом, разработанным в предыдущем модуле. Производится оценка модели на секретной выборке.

В этом же модуле участник готовит пакет документов по результатам работы, а также итоговую презентацию и доклад, требования к которым предоставляются в день С3.

Примерные этапы задания в модуле Е:

- 1) Оценка моделей на секретной выборке;
- 2) Программная документация, руководство для пользователя;
- 3) Итоговая презентация и доклад.

5. КРИТЕРИИ ОЦЕНКИ

Nº п/п	Раздел Матрицы навыков	Важность раздела
1	Организация и управление работой, коммуникация	6
2	Подготовка данных для анализа	18
3	Анализ и обработка данных	22
4	Построение моделей машинного обучения	15
5	Искусственный интеллект (ИИ) и машинное обучение	9
6	Разработка прикладных решений и документирование	30

Критерий оценки (Модули конкурсного задания)							Итого баллов за раздел Матрицы навыков
Разделы Матрицы навыков		A	B	C	D	E	
	1	2	2	0	0	2	6
	2	12	2	0	2	2	18
	3	2	18	0	0	2	22
	4	0	0	15	0	0	15
	5	0	0	9	0	0	9
	6	0	0	0	18	12	30
Итого баллов по Критерию оценки		16	22	24	20	18	100