Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра экономической информатики

Паспорт экзамена

по дисциплине «Анализ и визуализация бизнес-информации», 4 семестр

1. Методика оценки

Экзамен проводится в смешанной устно-письменной форме, по билетам. Билет состоит из 1 вопроса и 1 задачи и формируется по следующему правилу:

- первый вопрос выбирается из диапазона вопросов 1 – 21;
- второй вопрос – задача.

Таким образом, проверяются результаты обучения по дисциплине, соотнесенные с установленными в программе индикаторами достижения компетенций.

На экзамене преподаватель вправе задавать студенту уточняющие и дополнительные вопросы.

Форма экзаменационного билета

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Факультет ФБ

Билет № _____
к экзамену по дисциплине «Анализ и визуализация бизнес-информации»

---

1. Вопрос

   В чем состоит задача классификации данных в машинном обучении, дайте формальной определение и основные понятия задачи классификации? Опишите основной алгоритм классификации, с указанием вызываемых функций и классов. Приведите пример задач классификации. Перечислите основные методы классификации МО и дайте им краткую характеристику.

2. Задача.

Построить модель регрессии деревом решений на наборе данных «ENB2012_data». Построить график распределения данных. Выбрать признаки для построения модели, применив матрицу корреляции. Построить модель с различной вариацией параметров. Оценить модель метриками R^2, Mean Absolute Error,

Утверждаю: зав. кафедрой ____ _____ должность, ФИО
(подпись)

(дата)

2. Критерии оценки результатов обучения, соотнесенных с уровнями освоения индикаторов достижения компетенций

Ответ на экзаменационный билет засчитывается на **продвинутом** уровне, если студент проводит сравнительный комплексный анализ материала, выявляет проблемы, предлагает механизмы их решения, представляет количественные характеристики определенных процессов, приводит конкретные примеры, не допускает ошибок и способен обосновать выбор метода решения задачи. Совокупность результатов обучения по дисциплине (модулю) и соотнесенных с ними индикаторов достижения компетенций не содержит пробелов. Установленные в программе компетенции сформированы в полном объеме. Оценка составляет *от 35 до 40 баллов.*

Ответ на экзаменационный билет засчитывается на **базовом** уровне, если студент при ответе на вопросы формулирует основные понятия, дает характеристику процессов, явлений, проводит анализ причин, условий, способен представить количественные и качественные характеристики процессов, не допускает существенных ошибок при решении задачи. Совокупность результатов обучения по дисциплине (модулю) и соотнесенных с ними индикаторов достижения компетенций содержит несущественные пробелы. Установленные в программе компетенции сформированы на базовом уровне. Оценка составляет *от 27 до 34 баллов.*

Ответ на экзаменационный билет засчитывается на **пороговом** уровне, если студент при ответе на вопросы дает определение основных понятий, может показать причинно-следственные связи явлений, при решении задачи (если задача предусмотрена) допускает непринципиальные ошибки, например, вычислительные. Совокупность результатов обучения по дисциплине и соотнесенных с ними индикаторов достижения компетенций содержит пробелы. Установленные в программе компетенции сформированы на пороговом уровне. Оценка составляет *от* 20 *до* 26 *баллов.*

Ответ на экзаменационный билет считается **неудовлетворительным** (ниже порогового уровня), если студент при ответе на вопросы не дает определений основных понятий, не способен показать причинно-следственные связи явлений, при решении задачи допускает принципиальные ошибки. Совокупность результатов обучения по дисциплине (модулю) и соотнесенных с ними индикаторов достижения компетенций содержит существенные пробелы. Установленные в программе компетенции не сформированы. Оценка составляет *менее* 20 *баллов.*

3. Шкала оценки

Экзамен считается сданным, если сумма баллов по всем заданиям билета составляет от 20 до 40 баллов включительно. Сумма менее 20 баллов признается неудовлетворительным результатом промежуточной аттестации по дисциплине.

В общей оценке по дисциплине экзаменационные баллы учитываются в соответствии с правилами балльно-рейтинговой системы, установленными в НГТУ.

4. Вопросы к экзамену по дисциплине «Анализ и визуализация бизнес-информации»

1. Изобразите структурную схему элементов Data Science и искусственного интеллекта, дайте их понятия, охарактеризуйте основные составляющие, в контексте понятий и применений. Опишите основные отличия Data Science и искусственного интеллекта.

2. Опишите методы Data Science и последовательность их взаимодействия. Приведите пример применения методов Data Science.
3. Что называется Big Data, опишите основные характеристики больших данных, источники больших данных, способы хранения и обработки больших данных?
4. Что понимается под искусственным интеллектом (ИИ), виды ИИ, области применения ИИ, инструменты для создания систем с ИИ, приведите примеры систем с ИИ, кратко охарактеризуйте технологию разработки приведенных в примере систем?
5. Что понимается под машинным обучением (МО), опишите виды методов машинного обучения и их отличия друг от друга, перечислите фазы машинного обучения и дайте им краткую характеристику? В чем основное отличие алгоритмов машинного обучения от классических алгоритмов программирования? Приведите пример применения машинного обучения.
6. В чем отличие методов машинного обучения с учителем и без учителя? Дайте формализованное понятие обоих методов. Какие виды данных, в каком формате применяются они в каждом из подходов, в чем их отличие? Перечислите основные методы обучения с учителем и без учителя, дайте им краткую характеристику. Приведите пример задач, которые решаются с применение методов обучения с учителем и без учителя.
7. Что понимается под регрессионным анализом? Напишите общее уравнения простой и множественной линейной регрессии. Опишите основные отличия применения простой и множественной линейной регрессии. Опишите способ применения метода наименьших квадратов для подбора коэффициентов линейной регрессии. Опишите способы применения регрессии с помощью библиотек Pyhton для простой и множественной регрессии. Примеры задач МО с применением регрессии.
8. В чем состоит задача классификации данных в машинном обучении, дайте формальной определение и основные понятия задачи классификации? Опишите основной алгоритм классификации, с указанием вызываемых функций и классов. Приведите пример задач классификации. Перечислите основные методы классификации МО и дайте им краткую характеристику.
9. В чем заключается идея метода К-ближайших соседей, дайте формальное описание метода? Каким образом можно реализовать метод К-ближайших соседей с применением библиотек Pyhton, опишите основные параметры вызываемого класса? Приведите примеры задач МО с применением метода К-ближайших соседей.
10. В чем заключается идея метода наивного байесовского классификатора, опишите работу алгоритма с применением теоремы Байеса? Опишите работу алгоритма на простом примере. Каким образом можно реализовать метод наивного байесовского классификатора с применением библиотек Pyhton, опишите основные виды и параметры вызываемых классов? Приведите пример задач МО с применением метода наивного байесовского классификатора.
11. В чем заключается идея метода логистической регрессии, опишите работу алгоритма, на простом примере. Какие существуют требования к переменным при работе с логистической регрессией. Каким образом можно реализовать данный

метод с применений библиотек Pyhton, опишите основные виды и параметры вызываемого класса?

12. В чем заключается основная идея метода опорных векторов, опишите алгоритм на простом примере? Охарактеризуйте основные понятия метода, такие как опорные векторы, разделяющая гиперплоскость. Опишите принцип предсказания класса нового признака. Опишите принцип поиска лучшей гиперплоскости. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные виды и параметры вызываемого класса?
13. В чем заключается идея метода дерева решений, опишите алгоритм на простом примере. Объясните основные понятия метода, бинарное дерево, энтропия, прирост информации, отсечение дерева. Приведите пример задачи МО с применением метода дерева решений. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные параметры вызываемого класса? Опишите отличия дерева решений от случайного леса.
14. Что понимается под кластеризацией? В чем ее отличие от классификации? Опишите алгоритм кластеризации Kmeans и его отличия от метода Kmeans++, алгоритм поиска наилучшего количества кластеров. Перечислите и опишите основные метрики расстояний для данного алгоритма. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные виды и параметры вызываемых классов?
15. Назовите и опишите принципе работы иерархических методов кластеризации, меры расстояний между кластерами. Назовите и опишите основные особенности друих известных Вам методов кластеризации. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные параметры вызываемого класса?
16. Что понимается под ансамблевыми методами МО, их особенности и недостатки? Что такое беггинг, дайте определение и опишите алгоритм на простом примере? Что понимается под классификатором с голосованием? Приведите пример задач МО с применением беггинга. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные виды и параметры вызываемых классов?
17. Что понимается под ансамблевыми методами МО, их особенности и недостатки? Что такое бустинг, дайте определение и опишите алгоритм на простом примере? В чем отличие адаптивного и градиентного бустинга? Приведите пример задачи МО с применением бустинга. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные виды и параметры вызываемых классов?
18. Что понимается под ансамблевыми методами МО, их особенности и недостатки? Что такое стекинг, дайте определение и опишите алгоритм на простом примере? В чем отличие стекинга от других ансамблевых моделей? Приведите пример задачи МО с применением стекина. Каким образом можно реализовать данный метод с применений библиотек Pyhton, опишите основные виды и параметры вызываемых классов?
19. Что понимается под глубоким обучением? В чем отличие машинного от глубокого обучения? На простом примере опишите принцип реализации искусственного

нейрона, и опишите основные понятия, активационная функция, веса, смещение, входные значения, выходные значения. Что такое перцептрон? Опишите основные достоинства и недостатки применения глубокого обучения?

20. Какие виды метрик оценки моделей Вам известны? Чем обусловлен выбор метрики оценки модели? Опишите метрики, принципы их расчёта, а так же способ реализации с помощью библиотек Pyhton: Accuracy, Precision, Recall, F1-мера, ROC-AUC, Confusion Matrix.
21. Какие виды метрик оценки моделей Вам известны? Чем обусловлен выбор метрики оценки модели? Опишите метрики, принципы их рассчета, а так же способ реализации с помощью библиотек Pyhton: Mean Squared Error, R^2, Mean Absolute Error, ean Absolute Percentage Error, Symmetric Mean Absolute Percentage Error.

Задачи:

1. На основе набора данных banking.csv построить модель предсказания подписки клиента на срочный депозит, применив метод дерева решений. Перед построением модели произвести отбор признаков наиболее подходящих для моделирования, построить матрицу зависимости между выбранными признаками. Выполнить предподготовку данных. Произвести оценку важности признаков. Построить модель со стандартными параметрами и с применением методов настройки параметров. Оценить и сравнить качество обеих моделей F1-мерой.
2. На основе набора данных banking.csv построить модель предсказания подписки клиента на срочный депозит, применив метод К-ближайших соседей. Перед построением модели произвести отбор признаков наиболее подходящих для моделирования, построить матрицу зависимости между выбранными признаками. Выполнить предподготовку данных. Построить модель, настроить параметры модели. Оценить качество обеих моделей метриками Accuracy, Precision, Recall.
3. На основе набора данных banking.csv построить модель предсказания подписки клиента на срочный депозит, применив метод логистической регрессии. Перед построением модели произвести отбор признаков наиболее подходящих для моделирования. Выполнить предподготовку данных, в том числе избавить от дисбаланса данных . Произвести оценку важности признаков. Построить модель и настроить параметры модели. Оценить качество модели F1-мерой.
4. Построить модель предсказания о выдачи кредитов методом опорных векторов. В качестве набора данных файл Credit.csv. Для работы выделить признаки необходимые для работы модели, определить целевую переменную. Построить модель с различными вариациями параметров. Произвести оценку модели метриками Accuracy, Precision. Сравнить модель с моделью К-ближайших соседей.
5. Построить модель множественной линейной регрессии на наборе данных «ENB2012_data». Построить график распределения данных. Выбрать признаки для построения модели, применив матрицу корреляции. Построить модель с различной вариацией параметров. Оценить модель метриками Mean Squared Error, R^2.
6. Построить модель регрессии деревом решений на наборе данных «ENB2012_data». Построить график распределения данных. Выбрать признаки для построения модели, применив матрицу корреляции. Построить модель с различной вариацией параметров. Оценить модель метриками R^2, Mean Absolute Error,
7. Построить модели с применением наивного байесовского классификатора на наборе данных load_iris. Построить график распределения данных. Построить модель GaussianNB, MultinomialNB, ComplementNB. Сравнить оценки результаты работы моделей метриками Confusion Matrix, Precision, Recall