Оценочные средства для первой части курса. Примерный список вопросов для экзамена.

1. Определение модели бинарного выбора (логит и пробит регрессия)
2. Метод оценивания моделей конечного выбора.
3. Оценка качества моделей конечного выбора.
4. Построение дерева классификации с использованием алгоритма C4.5
5. Построение дерева классификации в рамках технологии CART.
6. Методы иерархической кластеризации. Метрики в пространстве наблюдений и методы агрегации данных.
7. Алгоритм кластерного анализа к-средних
8. Методы оценки качества кластеризации
9. Понятие нейронной сети. Основные элементы этой модели. Процесс обучения.

Оценочные средства для второй части курса. Примерный перечень вопросов к экзамену по всему курсу для самопроверки студентов:

1. Состав и назначение платформы обработки больших данных от IBM или другого крупного вендора.
2. Назначение и особенности программно-аппаратных комплексов Pure Data for Analytics и IDAA, отличия от традиционного подхода.
3. Использование IBM Watson Content Analytics и его место в информационно-аналитических системах.
4. Pure Data for Analytics: принципы работы, сжатие, ускорение, способы подключения, возможности аналитического блока.
5. Способы автоматизированной обработки естественного языка: назначение, существующие программные средства их описание и настройка.
6. Типы информационно-поисковых систем, информационно-поисковые языки. Фасет и фасетная классификация.
7. Архитектура HDFS, плюсы и минусы, блоки, операции записи/чтения, репликация, отказоустойчивость.
8. Подход NoSQL, особенности и сферы применения, сравнение с РСУБД
9. Hadoop - парадигмы MapReduce и Spark.
10. Описание подходов Pig&Hive. Wordcount на pig&hive
11. Место Hadoop в корпоративной ИТ-экосистеме
12. Обработка данных в режиме реального времени - основные принципы
13. Понятия потоков, кортежей, операторов
14. Streams, Storm, SPARK: краткое описание и особенности
15. Язык SPL и обработка данных скользящим окном
16. Новая парадигма работы с данными
17. Компетенции современного специалиста по данным
18. Что такое экспертная система
19. Превращение данных в информацию. Знания. Метаданные.
20. Источники и проблемы данных