**БАЛТИЙСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ "ВОЕНМЕХ"**

**им. Д. Ф. УСТИНОВА**

***Кафедра И9***

**ОТЧЕТ ПО НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ**

магистранта Николаева Антона Андреевича

группы И9М31

**ПРЕПОДАВАТЕЛЬ**

Верхолат А.М. / \_\_\_\_\_\_\_\_\_\_\_\_\_\_ /

Фамилия И.О. Подпись

“\_\_\_" \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ***2018 г.***

САНКТ-ПЕТЕРБУРГ

В этом семестре выполнялась работа по научно-исследовательской деятельности согласно составленному плану. Мою были решены некоторые задачи такие как

1. Анализ технологий для обработки неструктурированных данных.

Изучены различные технологии обработки неструктурированной информации, такие как Apache Hadoop и различные программы её экосистемы, Apache Spark и NoSQL-базы данных и их возможности для обработки неструктурированной информации, различные подходы и методы. Из изученных технологий выбрана оптимальная для построения тестового проекта обработки неструктурированных данных. Так же изучены некоторые смежные темы основной темы магистерской диссертации такие как Big Data (большие данные) и аналитика неструктурированных данных. Отдельно изучены подходы к анализу данных, такие как многомерный анализ, регрессия, классификация, кластеризация, поиск закономерностей A/B тестирование, алгоритм k-средних и другие подходы анализа неструктурированных данных, а также изучено специализированное ПО предлагающее реализацию этих методов. Так же были выбраны предполагаемые изучаемые данные (социальная сеть Твиттер). После изученного материала я перешел к построению проекта.

1. Выбор технологии создания тестового проекта.

Для тестового проекта выбрана платформа Cloudera Manager которая является дистрибутивом Apache Hadoop, включающая в себя компоненты её экосистемы такие как Apache Hive, Apache Zeppelin, Apache Sqoop.Сама платформа имеет возможности добавления других программ в экосистему, которые возможно понадобятся в ближайшем будущем. Платформы вторичной обработки данных еще не выбрана.

1. Определение темы научной диссертации, подготовка материалов для написания диссертации.

Так же в этом семестре определена окончательная тема диссертации и написано задание на диссертацию. Подготовлены некоторые материалы.