

## ОПЫТ ПРЕПОДАВАНИЯ ДИСЦИПЛИНЫ АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ В ВУЗЕ



**М.А. Амелин**

*Ассистент кафедры инженерной психологии и эргономики БГУИР, магистр экономических наук*

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь  
E-mail: michael.amelin@gmail.com*

*Abstract.* recent years, the use of machine learning methods has significantly increased and therefore research in this field is becoming more and more important. Competitiveness and success of organizations in the global economy will depend on the ability to use the methods of machine learning. Using of these methods will help to optimize those business processes in which it is necessary to analyze big data and automate administrative decisions.

В последние годы, использование методов машинного обучения значительно возросло и поэтому исследования в этой области становятся всё более важными. От умений качественно применять методы машинного обучения будут зависеть конкурентоспособность и успешность предприятий в глобальной экономике. С помощью методов машинного обучения предприятия получают возможность оптимизировать те свои процессы, в которых необходимо анализировать большие данные и автоматизировать управленческие решения.

В ходе подготовки к занятиям в 2016 / 2017 учебном году по дисциплине «Алгоритмы машинного обучения» автором статьи было пройдено обучение, сданы экзамены и получены сертификаты по следующим курсам, связанным с контекстом больших данных и алгоритмов машинного обучения:

- 1 Big Data Fundamentals, Armonk, New York, IBM (BDU), 2016;
- 2 Introduction to R-DataCamp, Armonk, New York, IBM (BDU), 2016;
- 3 Introduction to Machine Learning, Boston, DataCamp, 2016;
- 4 Intermediate R, Boston, DataCamp, 2016;
- 5 Intro to Statistics with R: Correlation and Linear Regression, Boston, DataCamp, 2016;
- 6 Hadoop Fundamentals I, Armonk, New York, IBM (BDU), 2016;
- 7 Big Data with SAP HANA Vora, Walldorf, SAP, 2016;
- 8 Enterprise Machine Learning in a Nutshell, Walldorf, SAP, 2016;
- 9 Imagine IoT, Walldorf, SAP, 2016;
- 10 Machine Learning for Data Science, San Francisco, Udemy, 2016;
- 11 Hands-on Industry 4.0, Potsdam, Hasso Plattner Institute for Software Systems Engineering (MOOC.House), 2016.

На рисунках 1–6 представлены некоторые из сертификатов, полученных во 2-й половине 2016 года.

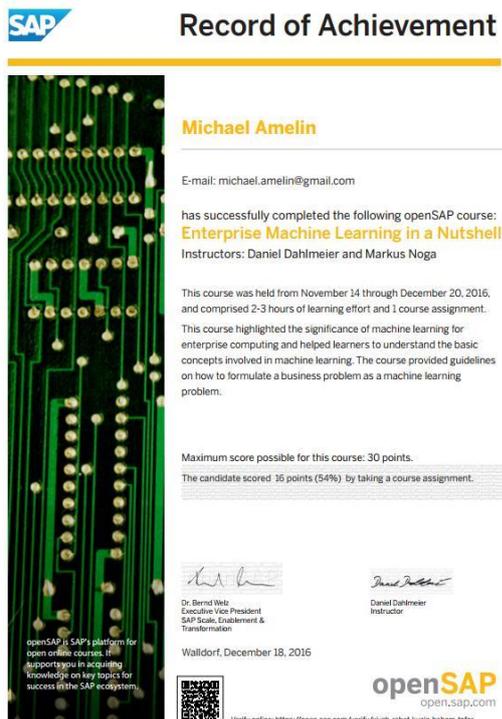


Рис. 1. Сертификат Enterprise Machine Learning in a Nutshell

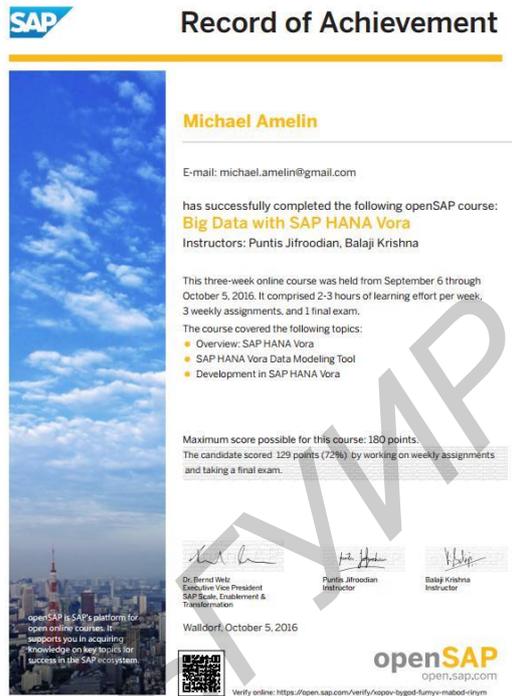


Рис. 2. Сертификат Big Data with SAP HANA Vora

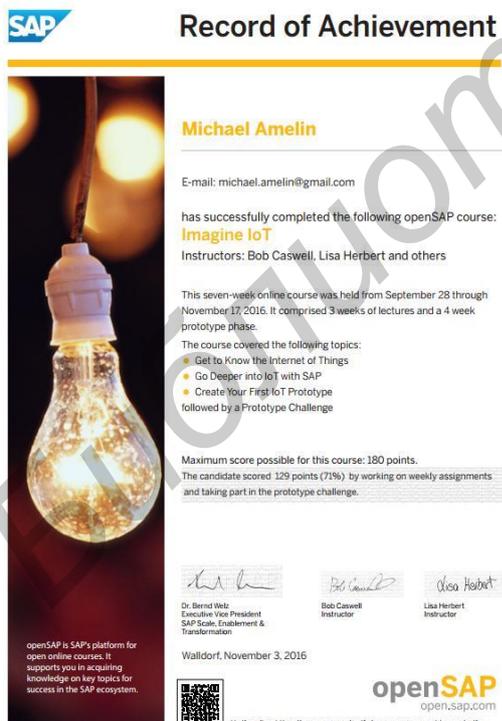


Рис. 3. Сертификат Imagine IoT

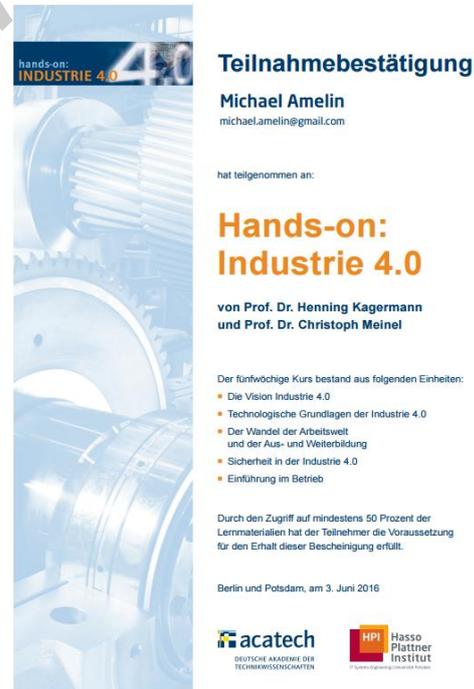


Рис. 4. Сертификат Hands-on: Industrie 4.0



Рис. 5. Сертификат Introduction to Machine Learning



Рис. 6. Сертификат Machine Learning for Data Science

Проводя экскурс в историю происхождения термина большие данные нужно упомянуть, что под ним подразумевается набор данных, который является настолько большим или сложным, что традиционные приложения обработки данных недостаточны для усвоения их. Также проблемы включают в себя анализ, поиск, совместное использование, хранение, передачу, визуализацию, обработку запросов, обновление и конфиденциальность информации. Термин «большие данные» часто относится просто к использованию прогностического анализа, аналитике поведения пользователей, и некоторым другим передовым методам анализа информации [1].

Организации могут повысить показатели своей эффективности и получить конкурентные преимущества при помощи надлежащей бизнес-аналитики. Существует четкая положительная корреляция между бизнес-анализом и предпринимательским успехом [2]. Увеличивающееся разнообразие типов данных, их источников, а также объема и скорости даёт организациям возможность для получения большего количества информации и последующего принятия обоснованных бизнес-решений.

Аналитика больших данных дополняет традиционные статические отчеты и помогает получить конкурентное преимущество организациям за счет верного предсказания бизнес-ситуации, оптимизации и улучшения адаптивности бизнес-модели [3]. Тем не менее, управление качеством данных становится всё более сложным, так как их разнообразие и количество источников постоянно увеличивается.

Большие данные имеют тенденцию постоянно расти в своём объеме, скорости и диапазоне типов и источников. При этом низкое качество данных является растущей проблемой. Наличие ошибок и несоответствий в источниках бизнес-данных и неправильная частота их сбора зачастую всё ещё не учитываются при проведении их аналитики.

Целью традиционной бизнес-аналитики является создание экономической ценности, своевременное нахождение существенных изменений в бизнес-процессах и принятие верных решений. Задачей же аналитики больших данных является извлечение полезной информации из массивных хранилищ данных. Такая информация может быть извлечена с помощью качественного сопоставления, в котором явление изучается путем наблюдения и выводов о переменных, которые измеряют сам феномен.

Описывая участие студентов в подготовке к сдаче экзамена по дисциплине «Введение в анализ данных с использованием технологии R» (Big Data University (BDU) от IBM), отметим, что они прошли онлайн-практику на электронной образовательной платформе DataCamp.



Рис. 7. Иллюстрации сертификатов студентов, прошедших онлайн-обучение на образовательной платформе DataCamp (BDU partner), Boston, USA

Также кроме прохождения курсов по «Основам BIGDATA», «Введение в анализ данных с использованием технологии R» (BDU) и «Введения в язык программирования R» (DataCamp) некоторые студенты смогли пройти следующие образовательные курсы на платформе Big Data University от IBM.

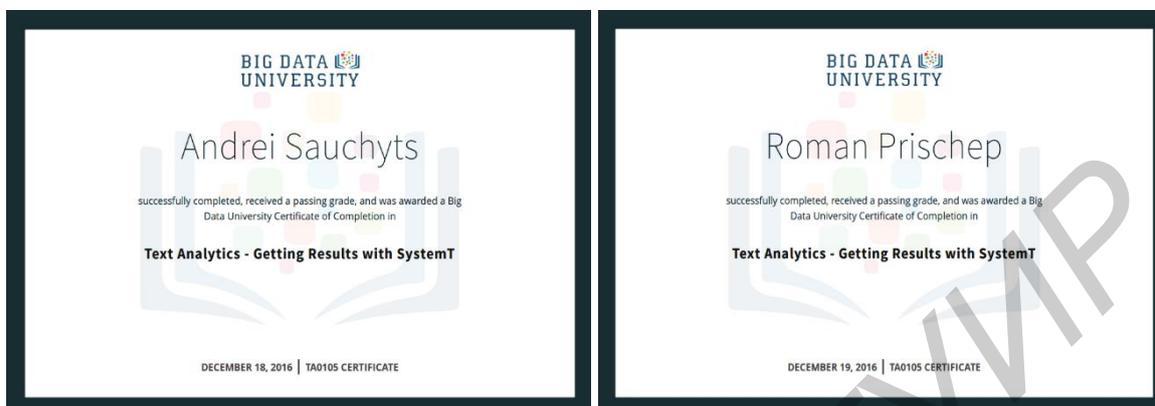


Рис. 8. Примеры сертификатов студентов, прошедших онлайн-обучение на образовательной платформе BDU

Далее рассмотрим некоторые вопросы, которые были предложены студентам для изучения разделов BIGDATA (в контексте алгоритмов машинного обучения) на лекционных, практических и лабораторных занятиях в БГУИР:

- Что такое Машинное обучение?
- Интеллектуальные приложения, работающие при помощи алгоритмов машинного обучения.
- Переход от бизнес-проблемы к задаче машинного обучения.
- Машинное обучение в контексте корпоративных вычислительных систем.
- Методы классификации в машинном обучении.
- Логистическая и линейная регрессия в машинном обучении.
- Примеры ИТ-приложений с учётом использования алгоритмов машинного обучения.
- Практические сферы применения алгоритмов машинного обучения.



Рис. 9. Слайды лекций «Алгоритмы машинного обучения»

В ходе следующих аудиторных занятий студентам также были предложены варианты расширения их знаний в области экосреды больших данных:





Приведём несколько примеров лабораторных занятий для студентов БГУИР по языку программирования R, в контексте подготовки к решению задач машинного обучения:



Рис. 16. Слайды лабораторных работ «Введение в язык R»



Рис. 17 Слайды лабораторных работ «Векторы, матрицы и списки в языке R»

В заключение добавим, что дальнейшая подготовка студентов на практических и лабораторных занятиях проводилась с учётом освоения следующих тем:

- Условные выражения и управление потоком данных;
- Логические операторы в языке R;
- Условные операторы;
- Цикл While;
- Цикл For;
- Введение в функции.

В результате курса «Алгоритмы машинного обучения» студенты готовы практически подходить к решению таких задач как: создание базовой модели прогнозирования, работе с матрицей неточностей, кластеризации видов, классификации и фильтрации спама в почте, моделирование будущих просмотров учётных записей в социальной сети LinkedIn и др.

#### Литература

- [1]. Амелин, М.А. Аналитика больших данных: вызовы и решения / М.А. Амелин, А.В. Артухевич // Topical areas of fundamental and applied research X: материалы X Международной науч.-практ. конф., NorthCharleston, 7–8 ноября 2016 г. : в 3 ч. / CreateSpace. – North Charleston, USA, 2016. Vol. 2. – С. 165–167.
- [2]. Maheshwari, A. Data Analytics Made Accessible / A. Maheshwari. – Seattle: Amazon Digital Services, 2014. – 156 p.
- [3]. Ankam, V. Big Data Analytics / V. Ankam. – Birmingham: Packt Publishing, 2016. – 326 p.