

*Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»  
им. В.И. Ульянова (Ленина), г. Санкт-Петербург, Россия*

***Аннотация.** В данной статье освещаются проблемы формирования знаний и навыков студентов в сфере машинного обучения. На текущий момент машинное обучение представляет собой обширную стремительно развивающуюся область науки, влияние которой сложно переоценить. Её изучение входит в программу подготовки специалистов в сфере информационных технологий, однако особенности профилей обучения, интересы и профессиональная направленность студентов учитываются далеко не всегда. В связи с этим в статье предлагается гибкий и структурированный подход к процессу подготовки IT-специалистов, отвечающий современным требованиям и тенденциям развития сферы информационных технологий.*

**Ключевые слова:** машинное обучение; образование; подготовка IT-специалистов; гибкий подход

**Введение.** В последнее время искусственный интеллект стал неотъемлемой частью нашей жизни. Его по праву можно назвать одним из самых актуальных и стремительно развивающихся направлений науки и техники. Интерес к этой области растёт с каждым днём, что способствует проникновению ИИ во множество сфер жизни [1]. Фундаментом для технологий искусственного интеллекта служит машинное обучение. В связи с этим всё острее становится необходимость соответствия образовательного процесса современным тенденциям развития информационных технологий в сфере машинного обучения и смежных с ним областей с целью подготовки высококвалифицированных IT-специалистов.

#### **Анализ текущего положения и описание проблематики.**

На сегодняшний день задачи, требующие применения технологий машинного обучения, решаются напрямую специалистами Data Science посредством написания программного кода, с помощью которого решаются все необходимые этапы обработки [2]:

- Исследование семантики и структуры данных, предоставляемых для обработки;
- Вычисление статистических значений данных и их атрибутов;
- Выполнение визуализации данных с целью их изучения;
- Предобработка данных, включающая их очистку, преобразование и подготовку для применения алгоритма;
- Построение модели и вычисление метрик её эффективности;
- Вычисление прогнозирующих значений с помощью модели на основе новых данных.

Подход, основанный на персональной работе специалиста с данными и программным кодом, позволяет достаточно эффективно решать небольшие и средние по сложности и размеру задачи. По этой причине «Машинное обучение» и смежные с ней дисциплины, входящие в программы подготовки IT-специалистов в высших учебных заведениях, направлены в первую очередь на формирование базовых навыков разработки программного кода в сфере машинного обучения, в то время как аналитической и управленческой составляющим данной области не уделяется достаточное внимание, что приводит к формированию узконаправленных знаний у студентов, а также усложняет процесс обучения дисциплине по программам подготовки специалистов в области анализа и управления.

На практике IT-компаниям всё чаще приходится сталкиваться с крупными и особо крупными проектами, в проектировании и реализации которых задействованы большие группы разработчики. Это значительно повышает сложность и трудоёмкость решения задач и требует от аналитического и управленческого звеньев более структурных и системных знаний в области Data Science, чем умение разрабатывать программный код.

По причинам, указанным выше, становится очевидной проблема нехватки формирования целостных предметно-ориентированных системных знаний студентов в области Data Science и потребность в корректировке и модификации соответствующих и смежных программ подготовки.

**Описание концепции решения проблемы.** В качестве решения обозначенной выше проблемы предлагается внедрение концепции, в основе которой лежит дифференциация дисциплин, направленных на формирование знаний и навыков в сфере машинного обучения, на два направления, предлагаемых студентам к изучению по выбору:

- «Разработка»;
- «Анализ и управление».

Теоретический материал каждого направления имеет общую и специальную части. Общая часть содержит информацию о концепции машинного обучения и принципах процесса обработки данных, описание используемых в данной области алгоритмов и методов, их особенностей, достоинств и недостатков, сведения о правилах построения моделей и метриках оценки их эффективности. В специальной теоретической части приводится информация, раскрывающая особенности конкретного направления и формирующая узкоспециализированные знания. При этом существенное различие заключается в подходах к организации практических занятий.

Практическая часть направления «Разработка» представляет собой текущий формат подготовки без каких-либо изменений. Содержанием указанной части является непосредственное решение студентом задач обработки данных с применением алгоритмов машинного обучения [3] путём разработки программного кода, включая:

- изучение языка Python;
- изучение библиотек, предназначенных для:
  - анализа и обработки данных. Примером реализации указанных методов является библиотека pandas;
  - применения алгоритмов и построения моделей. Например, данные возможности предоставляет библиотека scikit-learn [1];
  - визуализации данных двумерной и трёхмерной графикой. Примером такой библиотеки является Matplotlib.
- изучение специальных инструментов и средств разработки кода, таких как Jupyter Notebook, Яндекс DataSphere и других.

Данное направление ориентированно на формирование базовых навыков специалиста Data Science, способного применять индивидуальный подход к решению каждой поставленной задачи обработки данных, добиваться высокой степени эффективности анализа и качества его результатов.

Альтернативой текущему подходу к обучению и главным нововведением в соответствующие дисциплины является направление «Анализ и управление», содержащее в качестве специальной теоретической части информацию об особенностях проектирования, разработки и сопровождения систем, использующих технологии машинного обучения, а также специфики формирования и управления группами разработки соответствующих проектов.

В практической части указанного направления предлагается внедрение в процесс изучения дисциплин специального программного обеспечения, позволяющего осуществлять обработку данных и применение алгоритмов машинного обучения с помощью пользовательского интерфейса. Одной из таких систем является Orange Data Mining, разработанная университетом Любляны. Данный продукт содержит набор инструментов визуального программирования для интерактивной обработки и визуализации данных с применением машинного обучения [4]. Последовательность действий обработки («workflow») выстраивается пользователем с помощью размещаемых на рабочем поле и соединяемых между собой виджетов, реализующих различные настраиваемые операции преобразования и преобразования данных, алгоритмы построения моделей, методы визуализации и так далее. Подобный подход позволяет отойти от изучения конкретных реализаций алгоритмов и методов машинного обучения, предоставляемых множеством различных библиотек, сфокусировав внимание на самом процессе, его этапах и особенностях.

Направление «Анализ и управление» является особо ценным для программ подготовки IT-специалистов в области анализа и управления, а также для студентов, проявляющих в этом особый интерес, поскольку позволяет получить необходимое представление о сфере машинного обучения и особенностях управления в нём без лишней траты времени на погружение в программный код.

**Заключение.** Подводя итоги вышесказанному, хочется особо отметить потенциальный вклад от внедрения альтернативного направления в процесс преподавания дисциплины «Машинное обучение» и смежных с ней предметов. Реализация концепции выбора студентами наиболее интересующего их направления среди нескольких специализированных осуществит более гибкий и индивидуальный подход к обучению, что повысит качество подготовки будущих специалистов.

#### **Список литературы:**

1. Бевзенко С.А. Применение искусственного интеллекта и машинного обучения в разработке программного обеспечения // Инновации и инвестиции. 2023. №8.
2. Рашка С. Python и машинное обучение – Москва: ДМК Пресс, 2017. – 418 с. – ISBN 978-5-97060-409-0
3. В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святлов. Теория и практика машинного обучения: учебное пособие – Ульяновск: УлГТУ, 2017. – 290 с. ISBN 978-5-9795-1712-4.
4. Апельсиновый Data Mining // vc.ru // URL: <https://vc.ru/dev/198641-apelsinovyy-data-mining> (дата обращения: 20.03.2024). – Текст: электронный.

E. Yu. Saukhin, M. V. Livshits, B. F. Kachaev, P. P. Sharueva

Problems of formation of students' knowledge and skills in the field of machine learning

*Saint Petersburg Electrotechnical University, Russia*

**Abstract.** *This article highlights the problems of forming students' knowledge and skills in the field of machine learning. At the moment, machine learning is a vast rapidly developing field of science, the impact of which is difficult to overestimate. Its study is included in the training program for specialists in the field of information technology, however, the features of the study profiles, interests and professional orientation of students are not always taken into account. In this regard, the article proposes a flexible and structured approach to the process of training IT specialists that meets modern requirements and trends in the development of information technology.*

**Keywords:** machine learning; education; training of IT specialists; flexible approach