

Рецензия

на учебно-методическое пособие доцента ТПУ к.ф.-м.н. Губина Е.И.

«Methodology of big data preparation for predictive analysis»

(«Методология подготовки больших данных
для прогнозного анализа»)

Данное рецензируемое учебно-методическое пособие автор разработал в рамках курса «Введение в большие данные» для образовательной магистерской программы ТПУ 09.04.04 Программная инженерия («Технологии больших данных»/ «Big Data Solutions»). Следует особо подчеркнуть, что представленное учебно-методическое пособие предназначено для проведения занятий на английском языке, что является обязательным, т.к. данная магистерская программа англоязычная, что несет определенные особенности подачи материала.

Учебно-методическое пособие имеет понятную структуру, состоящую из следующих разделов: введение в большие данные, описательная статистика исходных данных, основные этапы подготовки исходных данных и выбор целевой функции, тренировочной и валидационной выборок, использование предсказательных моделей (на примере логистической регрессии), краткие выводы и заключение. В работе приведены источники и ссылки на видеоматериалы, которые помогают в освоении предложенного курса.

Считаю, что это учебно-методическое пособие «Методология подготовки больших данных для прогнозного анализа» соответствует курсу «Введение в большие данные» образовательной программы 09.04.04 «Технологии больших данных» и позволит студентам приобрести базовые знания и навыки при работе с большими данными в разных областях, включая финансовую индустрию, ритейл, инженерные приложения, маркетинг, медицину и другие.

Представленное учебно-методическое пособие может быть использовано как руководство при подготовке исходных («сырых») данных для дальнейшего прогнозного анализа. Учитывая, что сбор и подготовка данных является достаточно трудоемким и затратным мероприятием и первоначальные данные часто носят неструктурированный характер, то приведенные автором инструменты подготовки данных могут быть при необходимости еще дополнены.

Преподаватель Губин Е.И. представил в методической разработке базовые подходы к сбору и подготовке исходных данных. Предложенные автором

современные программные и статистические методики позволяют значительно сократить временные и финансовые затраты на этапе подготовки исходных данных и повысить эффективность и качество прогнозных результатов.

Особо хочется отметить, что в качестве современного программного инструмента для анализа данных автор использовал ПО «SAS», включая хорошо известный среди профессионалов Data Scientist алгоритмический подход «SEMMA». Это позволяет существенно повысить значимость и валидность предложенных подходов. На всех этапах подготовки данных, включая обнаружение выбросов («outliers»), пропущенных значений («missing»), преобразование переменных, поиск мультиколлинеарности и т.п. приведены конкретные примеры кода программ в «SAS-кодировке», что особенно ценно при действительно больших объемах данных.

Отличительной особенностью учебно-методического материала является его наглядность и желание автора продемонстрировать материал на конкретных примерах из реального бизнеса.

По мнению рецензента, данное пособие содержательное, последовательное и прекрасно наполненное руководство, которое могут использовать преподаватели ВУЗов, для подготовки специалистов в области больших данных в рамках ФГОС3++.

В заключение можно утверждать, что данное учебно-методическое пособие будет полезным не только студентам, но и преподавателям, специализирующимся при работе с большими данными.

Рецензент

Петровский Михаил Игоревич, к.ф.-м.н., РМР

Менеджер Академической Программы САС Россия/СНГ

Дата: « 20 » февраля 2020

Подпись: _____

