

Отзыв на магистерскую диссертацию  
студентки 2 курса дневного отделения кафедры теории управления  
Шелест Арины Александровны  
«О применении методов обнаружения выбросов  
к задаче исследования проб нефти»

Работа над магистерской диссертацией Шелест Арины Александровны выполнялась в рамках совместного проекта «Газпром» и ресурсного центра СПбГУ по направлению «химия». К сожалению, работа над данным проектом была прервана по независящим от математиков причинам, которые успели поучаствовать только на этапе предпроектного анализа данных. Тем не менее, проблема, затронутая в исследовании, носит крайне актуальный характер и потенциально может применена для схожих задач.

Магистерская диссертация Шелест А.А. посвящена статистическому анализу данных газовой хроматографии двадцати трех нефтяных проб, полученных из разных пластов Вынгапуровского месторождения (данные предоставлены ресурсным центром СПбГУ).

Основной целью работы являлось определение выбросов в данных, поступающих от хроматографов, с целью дальнейшего цензурирования выборки. При этом данные не согласовывались с нормальным законом распределения и большинство стандартных методов для цензурирования выборки в данной задаче не могли быть использованы. Обнаружение выбросов измерений является важной задачей, поскольку дальнейшая кластеризация данных по пластам крайне чувствительна к их наличию в выборочных данных. Кластеризация, в свою очередь, позволяет определить основные характеристики проб в различных пластах с тем чтобы в дальнейшем по пробам из неизвестных пластов определять их принадлежность к конкретному пласту. Данная проблема является актуальной прикладной задачей, имеющей важность для нефтехимической промышленности. Новизна определяется отсутствием публикаций по данной теме.

Глава I посвящена химической и геологической составляющей работы, содержит основные понятия из предметной области, необходимые для детального исследования задачи. Данная глава является реферативной частью, самостоятельно написанной Шелест А.А. на основе нескольких источников.

В главе II описывается результат статистической обработки данных, полученных из газового хроматографа. Приведены результаты дескриптивной статистики, произведена проверка данных на нормальность распределения несколькими методами, нормализация данных и корреляционный анализ. Для обнаружения выбросов на основе экспериментов была произведена модификация метода Тьюки.

В Главе III описан метод кластеризации k-means, приведены результаты его применения с различными способами снижения размерности. Для оценки описанных алгоритмов кластеризации введена метрика. Итогом работы является разбиение проб на кластеры и анализ полученных результатов.

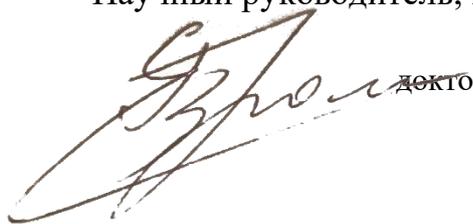
Высокая квалификация Шелест А.А. подтверждается не только глубоким погружением в предметную область и грамотным сопоставлением

нефтехимической задачи формальным математическим методам, но и программированием на высокоуровневом языке Python.

Выпускная квалификационная работа состоит из трех глав и двух приложений с фрагментом исходных данных и кодом программы на языке Python. Общий объем работы с приложением составляет 50 страниц. Список литературы содержит 17 наименований.

Считаю, что работа Шелест А.А. является законченным актуальным исследованием прикладного характера и заслуживает оценки «отлично».

Научный руководитель, профессор кафедры теории игр  
и статистических решений,  
доктор физико-математических наук



Громова Е.В.

18.05.2020