

Отзыв научного руководителя на выпускную квалификационную работу
аспиранта Санкт-Петербургского Государственного Университета
направления 09.06.01 «Информатика и вычислительная техника»

Пономарева А.А.

Тема «Сегментация пользователей мобильных операторов с помощью
моделей Больших Данных»

Данная выпускная работа Пономарева А. посвящена теме, которая только набирает популярность в современной научной теоретической и практической деятельности в России — Большим Данным. За рубежом тематика исследования получила распространение в последнее десятилетие, и ее обзор в зарубежной литературе представлен на порядок лучше, чем в отечественной. В работе А.Пономарева ставятся две задачи, имеющие прикладное значение в коммерческой деятельности оператора мобильной связи — задача по нахождению склонных к оттоку абонентов и задача построения маршрутов движения клиентов для оптимизации рекламных расходов. Если по первой задаче решения в научной литературе представлены достаточно хорошо, то по второй задаче практически нет научных работ ни в зарубежной, ни в отечественной практике.

Для написания работы был использован широкий набор источников литературы, касающихся деятельности азиатских мобильных операторов связи и банковских структур, а также был в достаточной мере проанализирован материал западноевропейских компаний. Обзор данной литературы касался в целом изучения методов кластеризации данных различными способами для нахождения оптимального варианта решения задачи, связанной с оттоком клиентов, а также поиска оптимального набора исследуемых абонентских характеристик. Обзор литературы, касающийся второй задачи по кластеризации маршрутов клиентов был сделан на основе нескольких работ финских ученых. Собранный материал из научных работ и научной литературы для обеих задач представляется достаточным для того, чтобы на практике применять способы кластеризации больших данных,

имеющихся в распоряжении информационных систем оператора, а также корректно интерпретировать полученные результаты. Теоретико-методологическая база работы опирается на несколько методов кластеризации, таких как random forest, нейросети, логистическая регрессия. Программное обеспечение кластеризации обеспечено на языке программирования Python.

Работа А.Пономарева носит не только интересный теоретический характер. В работе представлены результаты практического применения кластеризации на фактических реальных данных о клиентах оператора, а результаты работы приняты оператором в качестве базы для коммерческого использования. Необходимо отметить особым образом новизну решения задачи по кластеризации маршрутов, в котором удачным образом автор совместил теоретические наработки, имеющиеся в наличии в литературе, собственные предположения и идеи о характеристиках абонентов и связке данных по трафику, который регистрируется на базовых станциях, а также имеющийся в свободном доступе API компании Google для построения маршрутов.

Цель и задачи исследования сформулированы четко и в основном их достижение в работе можно считать успешным. Предположения относительно клиентских данных, выносимые в работе, сформулированы и часть из них доказана, а часть опровергнута самим автором. Структура работы представляется логичной. А,Пономарев является автором двух статей и соавтором с научным руководителем еще одной статьи, опубликованных в научных изданиях, рекомендованных ВАК России. Считаю, что работа Пономарева соответствует всем требованиям, предъявляемым к выпускным квалификационным работам и заслуживает оценки «отлично».

Научный руководитель

Заведующий кафедрой Системного программирования СПбГУ

Доктор физ.-мат. наук, профессор  Терехов А.Н.