О Т З Ы В

на бакалаврскую работу студентки 13.Б09-мм (461) группы кафедры информатики

математико-механического факультета СПбГУ

**Веревкиной Елены Борисовны**

«**Разработка программной системы извлечения и анализа неструктурированной информации по электронным торгам**»

В рамках данной работы была поставлена задача построения системы, которая позволит агрегировать информацию об автомобильной технике, реализуемой на электронных торгах по банкротству. На основе анализа собранной информации необходимо составить прогноз итоговой цены еще не реализованных лотов, а также получить статистику результатов торгов по завершившимся торгам.

В условиях экономического кризиса все больше физических и юридических лиц становятся банкротами, а их имущество чаще всего продается именно на электронных торгах, поэтому задача извлечения и анализа информации по электронным торгам несомненно является актуальной. Прогнозирование цены, по которой будет реализован конкретный автомобиль, поможет участникам соразмерить свои возможности и полученную премию, а затем сфокусироваться на наиболее перспективных лотах. Результаты анализа данных по уже завершившимся торгам могут помочь участникам будущих торгов выработать выигрышную стратегию в зависимости от своих целей.

Теоретическая часть работы включает в себя обзор аналогичных платформ, делается вывод об отсутствии функциональности прогнозирования. Затем были выявлены факторы, оказывающие наибольшее влияние на стоимость лота на электронных торгах, а также применен метод линейной регрессии для прогнозирования итоговой цены лота.

В качестве практической части дипломной работы автором была реализована программная система, включающая в себя 4 модуля: модуль извлечения неструктурированной информации из файлов форматов pdf, doc, docx, zip и извлечение слабоструктурированной информации из html страниц, а также запись полученной информации в базу данных; модуль, который производит обработку данных, их подготовку к применению в модели линейной регрессии; модуль, в котором осуществляется обучение, тестирование и применение модели линейной регрессии для прогнозирования итоговой цены лота; модуль, отвечающий за функционирование площадки для экспериментов, которая призвана помочь пользователю выработать собственную стратегию ведения торгов.

Все поставленные в рамках данной бакалаврской работы цели были достигнуты. В качестве недостатков можно отметить отсутствие представления собранной информации в виде графиков, а также отсутствие автоматизированного поиска наиболее выгодных для конкретного пользователя предложений.

Отмеченные недостатки существенно не снижают качество исследований и не влияют на главные теоретические и практические результаты диплома. Поэтому Веревкина Е.Б. может быть рекомендована к продолжению своей научной деятельности в магистратуре. Считаю, что её дипломная работа заслуживает оценки «ОТЛИЧНО».

Научный руководитель, к.ф.-м. н.,

доцент кафедры информатики

математико-механического факультета СПбГУ Григорьев Д.А.

29.05.2017