

Рецензия на выпускную квалификационную работу
Голованова Егора Олеговича
«Применение методов машинного обучения в области обработки естественного языка»

Выпускная квалификационная работа Голованова Е. О. посвящена вопросам бинарной классификации документов Интернета. Эта задача была поставлена ему, по утверждению автора ВКР, компанией Insecurityinsight, занимающейся мониторингом правовых нарушений по отношению к медицинским работникам.

В работе приводится схема решения задачи с предварительной обработкой текста документа приведением слов к их основе, удалением стоп-слов и выделение признаков (отображение документа на вектор n -мерного пространства признаков).

Перечисляются используемые в задачах классификации алгоритмы и обсуждаются методы оценки результатов. Приводятся общие описания (без конкретизации параметров для поставленной задачи) и результаты работы алгоритмов: наивного с различным построением выпуклой оболочки в пространстве признаков, k ближайших соседей, SVM, наивного байесовского классификатора.

Общее впечатление о работе: ее можно охарактеризовать как совокупность нескольких курсовых работ, но не как единого полноценного исследования поставленной задачи. В результате ознакомления с выпускной квалификационной работой следует сделать замечания:

1. На стр. 10 автор пишет «...с помощью *простого парсера*, были извлечены все, предоставляемые этим сайтом криминальные статьи». Вопрос: а зачем тогда все приводимые в ВКР исследования решаемой задачи, если она уже решена, и по утверждению автора использованием простого (какого?) парсера?

2. Не описано пространство признаков для решаемой задачи классификации.

3. Для алгоритма k ближайших соседей не объяснено выбранное количество соседей $k=5$, а также, почему автор выбрал именно те два расстояния, которые он исследовал.

4. При реализации алгоритма машины опорных векторов (SVM) не ясны параметры функций ядра.

5. Утверждается, что все алгоритмы реализованы автором на языке Python, однако в работе не приведено ни одного куска кода.

С учетом сделанных замечаний оцениваю дипломную работу оценкой **удовлетворительно**, но в случае успешной защиты оценка может быть повышена.

Рецензент, профессор кафедры
информационных систем,

доктор физико-математических наук



Матросов А.В.