

## ИССЛЕДОВАНИЕ ПОВЕДЕНЧЕСКИХ ХАРАКТЕРИСТИК ПОЛЬЗОВАТЕЛЕЙ НА ОСНОВЕ КЛАВИАТУРНОГО ПОЧЕРКА

**Цыганов М. А., Ямченко Ю. И.**  
(ООО “Лаборатория InfoWatch”)

В данной работе было произведено выделение признаков, создание и оценка качества модели для решения задачи аутентификации идентификации пользователя в информационной системе на основе клавиатурного почерка с использованием наиболее известных методов машинного обучения.

На начальном этапе решения поставленной задачи было произведено выделение признаков из логов пользователей, снятых при работе на клавиатуре в системе управления проектами JIRA. Информация в логах включала следующие данные: время события нажатия или отпускания клавиши, код клавиши, название клавиши, тип события (нажатие, отпускание).

На основании данных выгруженных логов были сформированы следующие классы признаков:

1. Учитывающие время удержания клавиш. Признак представляет собой разницу между временем нажатия и отпускания клавиши.
2. Учитывающие частоты нажатий.
3. Учитывающие характеристики биграмм и триграмм.
4. Обобщенные характеристики текста.

Для каждого признака, кроме частот появления в наборе пользователя специальных клавиш, было посчитано среднее по выборке и среднеквадратичное отклонение с учетом того факта, что пользователь может делать большие паузы при наборе текста. Общее число признаков равно 154.

На основании выделенных признаков по данным клавиатурного набора для каждого из пользователей были построены векторы признаков. Затем был проведен факторный анализ, по результатам которого было произведено выделение малознача-

щих признаков и сокращение размерности исходных векторов признаков.

Следующая стадия работы заключалась в построении моделей классификаторов и оценке качества их работы. Для построения моделей классификаторов использовался ряд известных методов обучения классификаторов, предоставляемых Python библиотекой scikit-learn, включая алгоритмы KNN, SVM и Random Forest.

Также была произведена оценка точности и скорости работы построенных моделей классификаторов на тестовой выборке логов пользователей. Оценка точности работы классификаторов производилась по критериям: AUC ROC, AUC PR, критерий F1, accuracy, FRR, FAR.

По результатам анализа значений приведенных критериев для каждой из моделей была выбрана оптимальная модель классификатора для решения задач идентификации и аутентификации пользователей.

### **Литература**

ХОДАШИНСКИЙ И.А., Савчук М.В., Горбунов И.В., Мещеряков Р.В. Технология усиленной аутентификации пользователей информационных процессов // Доклады ТУСУРа, №2, часть 3. - 2011. - 236-248 с.

ЯНДИЕВ И. Б. Исследование временных характеристик клавиатурного почерка для быстрой аутентификации личности // Молодой ученый. — 2017. — №14. — С. 154-158. — URL <https://moluch.ru/archive/148/41543/> (дата обращения: 10.04.2018).

ЯНДИЕВ И. Б. Исследование временных характеристик клавиатурного почерка для быстрой аутентификации личности // Молодой ученый. — 2017. — №14. — С. 154-158. — URL <https://moluch.ru/archive/148/41543/> (дата обращения: 10.04.2018).