

УДК 519.715

Сидор В. –ст. гр. СНм-61

Тернопільський національний технічний університет імені Івана Пулюя

ПОРІВНЯННЯ МЕТОДІВ NAIVE BAYES, MULTINOMIAL NAIVE BAYES ТА BERNOULLI NAIVE BAYES ДЛЯ ЗАДАЧІ КЛАСИФІКАЦІЇ LINKEDIN ПРОФІЛІВ

Науковий керівник: к.т.н., доцент Фриз М. Є.

Sydor V.

Ternopil Ivan Puluj National Technical University

COMPARISON OF METHODS NAIVE BAYES, MULTINOMIAL NAIVE BAYES AND BERNOULLI NAIVE BAYES FOR THE PROBLEM OF CLASSIFICATION OF LINKEDIN PROFILES

Supervisor: associate professor Fryz M. Ye.

Ключові слова: Naive Bayes, Порівняння, Класифікатор

Keywords: Naive Bayes, Comparison, Classifier

У доповіді буде розглянуто порівняння класифікаторів на основі алгоритму Naive Bayes та досліджено, яка модель із Naive Bayes класифікаторів найкраще проявить себе для класифікації тексту.

Задано однаковий навчальний набір для класифікації працівників в компанії. Компанія складається із 13 відділів: Founder, CEO, Corporation, Finance, Product Manager, Marketing, Sales, Support, IT, Partner ship, Owner та Other.

Навчальний набір налічує в собі 950 записів, по яких класифікатори навчатимуться. Для перевірки, вручну прокласифіковано 91 запис, по якому здійснюється порівняння.

Протестовано класифікатори на основі алгоритмів Naive Bayes, а саме: звичайний Naive Bayes, Multinomial Naive Bayes та Bernoulli Naive Bayes.

Перше порівняння - за точністю правильної відповіді класифікаторів. Найкращі результати показали Multinomial Naive Bayes - 74.5% та простий Naive Bayes - 72.3%, Bernoulli Naive Bayes показав набагато гірший результат - 60% правильності.

Другий етапом порівняння - за загальним часом роботи класифікатора. Multinomial Naive Bayes та Bernoulli Naive Bayes, обробляють 91 запит за 360 мс, класичний Naive Bayes - за 5.1 с.

Із проведених тестів за точністю класифікатора та часом класифікації, найкращі результати показав Multinomial Naive Bayes, який показав найменший час роботи класифікатора та найбільшу точність правильної відповіді. Також непогано показав класичний Naive Bayes, але при умові не великої кількості даних.

Література

1. Christopher D. Manning. Introduction to Information Retrieval / D. Manning Christopher - Cambridge University Press, 2013. – 581 с.
2. Gupta G. K., Introduction to Data Mining with Case Studies / G. K. Gupta Prentice Hall of India, New Delhi, 20016. – 200 с.