

Висновок.

Аналіз текстів є важливою задачею в сучасному інформаційному суспільстві. Розробка ефективних методів обробки текстів відіграє ключову роль у витягуванні корисної інформації з великого обсягу текстових даних. У даній статті ми розглянули види та класифікації текстів, а також основні етапи обробки текстів. Крім того, був наведений приклад використання нейромережі для обробки текстів. Опанування та вдосконалення методів обробки текстів допоможе в розвитку автоматичного аналізу текстів та покращенні різних застосувань, від інформаційного пошуку до машинного перекладу.

Список використаної літератури

1. *Natural Language Processing, NLP [Електронний ресурс]* – Режим доступу до ресурсу:
<https://evergreens.com.ua/ua/articles/natural-language-processing.html>.
2. *Автоматична обробка текстів природною мовою та комп'ютерна лінгвістика : книга, посібник.* / Є. Большакова, Е. Клишинський, Д. Ланде, А. Носков, О. Пескова та Є. Ягунова. – М. : МІЕМ, 2011. – 272 с
3. *Уоссермен Ф. Нейрокомп'ютерна техніка / Ф. Уоссермен.* – М. : Мир, 1992. – 238 с.

УДК 004.01

*Бойко У. В., Студент 1 курсу спеціальності 122 «Комп'ютерні науки» СО Магістр
Нескородєва Т. В., д.т.н., доцент кафедри інформаційних технологій*

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАТАСЕТУ «AUTO»

Донецький національний університет імені Василя Стуса, м. Вінниця

Вплив різних факторів на витрати палива автомобілів є актуальною та важливою проблемою у сучасному світі, де стала популярною енергоекономія та зменшення негативного впливу на довкілля.

В даній дослідженні було проведено аналіз набору даних Auto, що містить інформацію про різні характеристики автомобілів, з метою встановлення залежності між цими характеристиками та витратами палива.

Набір даних Auto складається з 392 спостережень за наступними 9 змінними:

- mpg – кількість пройдених миль на одному галоні палива;
- cylinders – кількість циліндрів у автомобіля;
- displacement – об'єм двигуна в кубічних дюймах;
- horsepower – потужність двигуна;
- weight – вага автомобіля;
- acceleration – час розгону від 0 до 60 миль/год в секундах;

- year – рік моделі автомобіля;
- origin – походження автомобіля (1. Американський, 2. Європейський 3. Японський);
- name – назва транспортного засобу.

Застосована методологія включала використання лінійної регресії, лінійний дискримінантний аналіз (LDA), квадратичний дискримінантний аналіз (QDA) та логістичної регресії для побудови моделі передбачення витрат палива. Було визначено найбільш тісно пов'язані змінні та проведено порівняння з іншими моделями передбачення.

Використана функція `best_predictor()`, яка на виході видає таблицю значень коефіцієнтів детермінації всіх вказаних змінних у наборі даних, від найкращого до найгіршого, тим самим показує які змінні є найкориснішими для передбачення `mpg`.

```
> best_predictor(select(Auto, -name), "mpg")
[1] "Response variable: mpg"
  varname vartype   R2 R2_log R2_quad max_R2
1  weight numeric 0.693 0.713  0.715 0.715
2 displacement numeric 0.648 0.686  0.689 0.689
3 horsepower numeric 0.606 0.668  0.688 0.688
4  cylinders numeric 0.605 0.603  0.607 0.607
5     year numeric 0.337 0.332  0.368 0.368
6   origin numeric 0.319 0.330  0.332 0.332
7 acceleration numeric 0.179 0.190  0.194 0.194
8      mpg numeric   NA    NA    NA   NA
```

Рисунок 1 – Таблиця ефективності предикторів для передбачення `mpg`

Результати дослідження підтвердили існування значимих залежностей між різними факторами та витратами палива. Основні фактори, що впливають на витрати палива, включають об'єм двигуна, вагу автомобіля, кількість циліндрів та його потужність.

Використання QDA, LDA, лінійної та логістичної регресії дозволило створити ефективну модель для передбачення витрат палива. Порівняння моделей показало, що всі вони є досить ефективними, маючи низьке значення похибки прогнозування витрат палива на контрольних та тестових даних. Однак квадратичний дискримінантний аналіз показав найнижче значення похибок, що робить його найкращим підходом для цього завдання.

```
> set.seed(2)
> qda_mpg <- train(mpg01 ~ cylinders + displacement + weight + horsepower,
+                 data = train,
+                 method = "qda",
+                 trControl = ctrl)
> 1 - qda_mpg$results$Accuracy
[1] 0.08991228
> predicted_qda <- predict(qda_mpg, newdata = test, type = "raw")
> mean(predicted_qda != test$mpg01)
[1] 0.1173469
```

Рисунок 2 – QDA-модель по навчальних даних для передбачення `mpg01`

Під час перехресної перевірки на навчальних даних, помилка складає близько 0.0899, що означає, що модель правильно передбачає приблизно 91.01% даних.

На контрольних даних, модель має помилку близько 0.1173, що означає, що вона правильно передбачає приблизно 88.27% даних.

Дослідження виявило потенціал для подальшого розширення та удосконалення методів передбачення витрат палива, включаючи врахування додаткових факторів та застосування більш складних моделей машинного навчання.

Результати дослідження можуть бути корисними для автомобільної промисловості, енергетичних компаній та політиків, що працюють у сфері сталого розвитку, спрямованого на зменшення споживання палива та негативного впливу на довкілля.

Висновки цього дослідження свідчать про важливість аналізу впливу різних факторів на витрати палива автомобілів та підкреслюють роль квадратичного дискримінантного аналізу у передбаченні цих витрат. Дані результати можуть сприяти розвитку більш ефективних та екологічно чистих автомобільних технологій.

Список літературних джерел

1. Wickham H. *R for Data Science* / H. Wickham, G. Grohl. – Sebastopol, California, USA: O'Reilly Media, 2017. – 491 с.
2. *Пакет caret – універсальний інтерфейс для доступу до десятків алгоритмів машинного навчання [Електронний ресурс] // Аналіз і візуалізація даних. – 2019. – Режим доступу до ресурсу: <https://r-analytics.blogspot.com/2015/06/caret.html>.*
3. *Введення в статистичне навчання з прикладами на мові R* / Г.Джеймс, А. Уїттон, Т. Хасті, Р. Тібшірані., 2017. – 456 с.

УДК 004.01

*Дужак А.О., студент 4 курсу,
спеціальності 122 «Комп'ютерні науки»
Ніколюк П.К., д.ф-м.н., професор, професор
кафедри інформаційних технологій*

РЕКОМЕНДАЦІЙНА СИСТЕМА ПІДБОРУ ЛІКІВ

Донецький національний університет імені Василя Стуса, м. Вінниця

Дана робота присвячена вивченню проблеми добору лікарських засобів для пацієнтів з різними характеристиками організму. У роботі було проведено аналіз літературних джерел, що дозволило визначити основні методи добору лікарських засобів, їх переваги та недоліки. Зокрема, було розглянуто традиційні методи добору лікарських засобів, такі як використання стандартів лікування та результатів клінічних досліджень.

Також було розглянуто новітні методи добору лікарських засобів, такі як персоналізована медицина, яка базується на індивідуальних характеристиках пацієнта та використанні молекулярних маркерів. Для впровадження цього