



یادگیری ماشین

پروژه نهایی NLP

عنوان پروژه: تحلیل احساسات در متن‌های رسانه‌های اجتماعی

مراحل انجام پروژه:

1. جمع‌آوری داده:

- برچسب گذاری روی 200 داده متنی. (مثبت : 0 ، خنثی : 1 ، منفی : 2)

2. پیش‌پردازش داده:

- بررسی وجود نویز.
- بررسی وجود داده گمشده.
- پاکسازی اطلاعات اضافه.
- توکن بندی داده‌ها¹.
- حذف کلمات پرتکرار².
- Stemming.

3. استخراج ویژگی:

- استفاده از روش‌های زیر برای استخراج ویژگی:

Bag-of-words ○

TF-IDF ○

4. انتخاب مدل:

- استفاده از ویژگی‌های استخراج شده و مدل‌های یادگیری ماشین ذکر شده برای تحلیل احساسات.

SVM ○

Naïve Bayes ○

¹ Tokenization

² Stop words

- ادغام یک مدل BERT از پیش‌آموزش دیده (BERT-based-uncased)³ بر اساس متن و استفاده از این مدل برای تخمین احساسات با مدل های زیر و ارزیابی آنها : (امتیازی)

○ SVM

○ Naïve Bayes

○ Logistic Regression

5. آموزش و ارزیابی مدل:

- تقسیم داده‌ها به مجموعه های تست و آموزش.
- سنجش مدل ها استفاده شده روی مجموعه آموزش و تست معیار هایی مانند Accuracy.
- تهیه Confusion Matrix برای هر مدل.

موارد تحویلی:

- تحویل کد پروژه.
 - تهیه گزارش دقیق از کارهای انجام شده و نتایج بدست آمده.
 - آمادگی برای ارائه پروژه (به صورت آنلاین و گروهی).
- توجه:** برای کسانی که تصمیم به انجام بخش امتیازی دارند، ارزیابی های جامع الزامی است.

موفق باشید.

³ <https://huggingface.co/bert-base-uncased>