

# Sentence Transformer

## Utilisation de la technologie BERT

### Genres

['Comedy', 'Action'...]

### Resume

Monsieur Hulot curiously wanders around a high-tech Paris, paralleling a trip with a group of American tourists. Meanwhile, a nightclub/restaurant prepares its opening night, but it's still under construction.

```
model = SentenceTransformer('all-MiniLM-L6-v2')

dataset['genre'] = ''
dataset['themaScore'] = 0.0

def compute_themaScore(dataset):
    data = dataset.copy()

    for i in range( len(data) ):
        bestScore = 0.0
        bestGenre = ''
        for j in range(len(data.iloc[i]['genres'])):
            modelGenre = model.encode(str(data.iloc[i]['genres'][j]))
            modelResume = model.encode(str(data.iloc[i]['resume']))
            result = util.cos_sim(modelGenre, modelResume).tolist()[0][0]
            if result > bestScore:
                bestScore = result
                bestGenre = data.iloc[i]['genres'][j]

        #data.loc[i, 'themaScore'] = bestScore
        #data.loc[i, 'genre'] = bestGenre
        data['themaScore'][i] = bestScore
        data['genre'][i] = bestGenre

        #print(i, '/', len(data), bestGenre, bestScore)
        #print(data.iloc[i])

    return data

t11 = compute_themaScore(dataset[290000:])
t11 = t11.drop(columns=['genres'])
t11.to_csv(r'temps/t11.csv', index = False, header=True)
```

Python

### Genre

Comedy

### ThemaScore

0.09567847



# Machine Learning