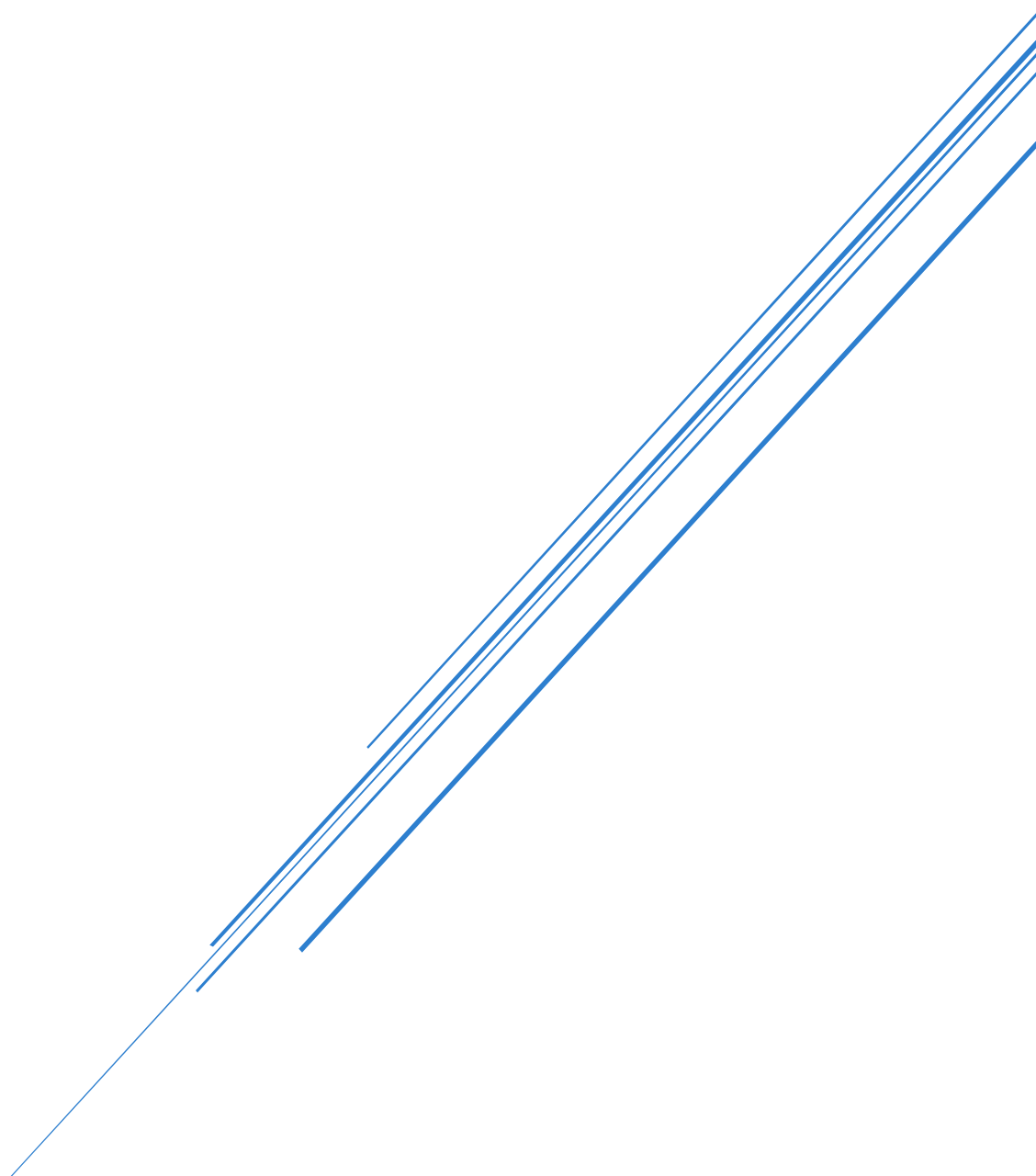


LIVECHAT ASSISTANCE

with
ParsBert



Erfan Arabioun

1. مقدمه

هدف از این پروژه، طراحی و پیاده‌سازی یک سیستم پاسخ‌دهی هوشمند است که با استفاده از یک پایگاه دانش از پیش آماده‌شده، به سؤالات کاربران پاسخ می‌دهد. این سیستم با بهره‌گیری از مدل‌های پیشرفته زبان طبیعی، از جمله ParsBERT، قادر است با تجزیه و تحلیل معنایی، نزدیک‌ترین پاسخ را برای هر سؤال ارائه دهد.

2. مفاهیم کلیدی

2.1 مدل زبان طبیعی (Natural Language Model)

مدل زبان طبیعی، مانند ParsBERT، الگوریتمی است که می‌تواند زبان انسانی را درک و پردازش کند. این مدل بر اساس شبکه‌های عصبی عمیق آموزش‌دیده و قادر به انجام وظایفی مانند طبقه‌بندی متن، تولید متن و درک معنای آن است.

2.2 توکنایزر (Tokenizer)

توکنایزر بخشی از مدل‌های NLP است که متن ورودی را به قطعات کوچک‌تر (توکن‌ها) تقسیم می‌کند. برای مثال، عبارت «چگونه حساب کاربری بسازم؟» به کلماتی مانند «چگونه»، «حساب»، «کاربری» و «بسازم» تجزیه می‌شود.

هدف: آماده‌سازی متن به گونه‌ای که مدل بتواند آن را پردازش کند.

2.3 امبدینگ (Embedding)

امبدینگ یک نمایش عددی از متن است که معنا و مفهوم آن را به صورت بردار در فضای چندبعدی نمایش می‌دهد. بردارها اجازه می‌دهند که مدل شباهت معنایی بین متون مختلف را محاسبه کند. مزیت: امکان مقایسه جملات مختلف بر اساس معنا، نه فقط کلمات.

2.4 شباهت کسینوسی (Cosine Similarity)

شباهت کسینوسی معیاری است که نشان می‌دهد دو بردار چقدر به یکدیگر نزدیک هستند. مقدار آن بین ۰ (کاملاً متفاوت) تا ۱ (کاملاً مشابه) است.

فرمول:

$$\text{Cosine Similarity} = \frac{A \cdot B}{||A|| \cdot ||B||}$$

کاربرد در پروژه: یافتن نزدیکترین سؤال پایگاه دانش به سؤال کاربر.

3. معماری سیستم

1. بارگذاری پایگاه دانش:

فایل Excel حاوی سؤالات و پاسخها به سیستم وارد می شود.

2. تبدیل سؤالات به بردار (Embedding):

سؤالات پایگاه دانش به بردارهای عددی تبدیل و ذخیره می شوند.

3. پیدا کردن پاسخ مناسب:

با استفاده از شباهت کسینوسی، پاسخ نزدیکترین سؤال به سؤال کاربر از پایگاه دانش بازیابی می شود.

4. کد و توضیحات خطبه خط

4.1 نصب کتابخانه ها

```
!pip install transformers torch scikit-learn pandas
```

- transformers: برای کار با مدل های زبان طبیعی مانند ParsBERT.
- torch: برای اجرای مدل های هوش مصنوعی با استفاده از PyTorch.
- scikit-learn: برای محاسبه شباهت بردارها.
- pandas: برای خواندن و پردازش فایل های Excel.

4.2 بارگذاری مدل و توکنایزر ParsBERT

```
from transformers import AutoTokenizer, AutoModel  
model_name = "HooshvareLab/bert-fa-base-uncased"  
tokenizer = AutoTokenizer.from_pretrained(model_name)
```

```
model = AutoModel.from_pretrained(model_name)
```

- `AutoTokenizer`: تبدیل متن به فرمت عددی قابل پردازش توسط مدل.
- `AutoModel`: بارگذاری مدل ParsBERT برای پردازش زبان فارسی.
- `from_pretrained`: دانلود مدل از مخزن آنلاین.

4.3 تولید بردارها (Embedding)

```
def encode_questions(questions):  
    embeddings = []  
    for question in questions:  
        inputs = tokenizer(question, return_tensors="pt", truncation=True, padding=True,  
max_length=128)  
        outputs = model(**inputs)  
        embeddings.append(outputs.last_hidden_state.mean(dim=1).detach().numpy())  
    return np.vstack(embeddings)
```

- `encode_questions`: تابعی برای تولید بردارهای معنایی.
- `tokenizer(question)`: تبدیل متن به توکن‌ها.
- `model(**inputs)`: تولید بردار معنایی با مدل ParsBERT.
- `mean(dim=1)`: میانگین‌گیری از بردارهای خروجی.
- `np.vstack`: ترکیب همه بردارها در یک ماتریس.

```
faq_embeddings = encode_questions(questions)  
np.save("faq_embeddings.npy", faq_embeddings)
```

- `faq_embeddings`: ماتریس بردارهای تمامی سؤالات پایگاه دانش.
- `np.save()`: ذخیره بردارها برای استفاده در آینده.

4.4 یافتن پاسخ مناسب

```
from sklearn.metrics.pairwise import cosine_similarity

def find_best_answer(user_question, questions, embeddings):
    user_embedding = encode_questions([user_question])[0]
    similarities = cosine_similarity([user_embedding], embeddings)
    best_match_idx = similarities.argmax()
    return questions[best_match_idx]
```

- `user_embedding`: تولید بردار سؤال کاربر.
- `cosine_similarity`: مقایسه بردار کاربر با بردارهای پایگاه دانش.
- `similarities.argmax()`: یافتن برادری که بیشترین شباهت را دارد.

```
user_question = "چگونه حساب کاربری بسازم؟"
best_answer = find_best_answer(user_question, faq_data['Answer'].tolist(), faq_embeddings)
print("Best Answer:", best_answer)
```

- `user_question`: سؤال نمونه کاربر.
- `best_answer`: نزدیک‌ترین پاسخ موجود در پایگاه دانش.

5. نتیجه‌گیری

این سیستم قابلیت:

1. پاسخ‌دهی به سؤالات کاربران به‌صورت هوشمند.
2. یادگیری و به‌روزرسانی پایگاه دانش.
3. پردازش سریع و دقیق سؤالات.