



تحلیل احساس کاربران نسبت به خدمات بانکی با استفاده از پردازش زبان طبیعی (NLP)

سروش عزیزی

سینا سپهوند

چکیده (Abstract)

در این پروژه، یک سیستم جامع برای تحلیل احساس نظرات کاربران ایرانی درباره اپلیکیشن‌های بانکی طراحی و پیاده‌سازی شده است. نظرات از فروشگاه کافه‌بازار جمع‌آوری شده و پس از پردازش متن فارسی، با استفاده از مدل‌های مختلف یادگیری ماشین و مدل‌های پیشرفته زبانی مانند ParsBERT تحلیل احساسات انجام شده است. این تحلیل به‌منظور درک بهتر تجربه کاربری و بهبود خدمات بانکی انجام گرفته و دقت مدل نهایی (Ensemble) به 86.3% رسیده است.

مقدمه

تحلیل احساسات یکی از کاربردی‌ترین حوزه‌های پردازش زبان طبیعی است که به سازمان‌ها امکان می‌دهد بازخوردهای کاربران را به‌صورت خودکار تحلیل و تفسیر کنند. در حوزه بانکداری، این تحلیل می‌تواند برای درک مشکلات کاربران، بهبود خدمات و افزایش رضایت مشتریان استفاده شود. هدف این پروژه، تحلیل نظرات کاربران فارسی درباره خدمات اپلیکیشن‌های بانکی ایران است.

جمع‌آوری داده‌ها

- منبع داده: کافه‌بازار (فروشگاه اپلیکیشن‌های ایرانی)
- نوع داده: نظرات کاربران درباره اپلیکیشن‌های رسمی بانک‌ها مانند بانک ملت، صادرات، پارسیان، ملی، تجارت، سامان
- روش جمع‌آوری: Scraper اختصاصی با نرخ محدودسازی درخواست‌ها
- حجم داده: بیش از 1000 کامنت فارسی
- امتیاز ویژه: داده‌ها به صورت اختصاصی گردآوری شده‌اند

پیش‌پردازش متن فارسی

با توجه به چالش‌های خاص زبان فارسی، مراحل زیر پیاده‌سازی شد:

- نرمال‌سازی نویسه‌ها (مثل "ی" عربی به فارسی)
- حذف کاراکترهای غیرضروری (URL، ایموجی، اعداد)
- حذف کلمات توقف فارسی
- تبدیل به فرم پایه (Stemming)
- توکن‌سازی با درنظر گرفتن ZWNJ (نیم‌فاصله)
- حداقل و حداکثر طول کامنت برای حذف نویز: بین 10 تا 500 کاراکتر

نمایش متن (Text Representation)

سه روش مختلف برای تبدیل متن به ویژگی‌ها استفاده شد:

- **TF-IDF**: برای مدل پایه
- **Word2Vec**: آموزش دیده بر دیتاست بانکی فارسی
- **ParsBERT Embeddings**: استفاده از مدل پیش‌آموزش دیده ParsBERT

مدل‌های یادگیری و پیاده‌سازی

مدل‌های مختلف با داده‌ها آموزش داده شده‌اند:

- **Logistic Regression + TF-IDF (Baseline)**
- **CNN / LSTM + Word2Vec**
- **Fine-tuned ParsBERT**
- **Ensemble (ترکیب مدل‌ها)**

ارزیابی مدل‌ها

Model	Accuracy	F1-Score	Precision	Recall
-----	-----	-----	-----	-----
Logistic + TF-IDF	0.724	0.689	0.705	0.694
CNN + Word2Vec	0.756	0.731	0.748	0.739
LSTM + Word2Vec	0.768	0.742	0.751	0.745
ParsBERT (Fine-tuned)	0.847	0.823	0.831	0.829
Ensemble (All Models)	0.863	0.841	0.849	0.844

تحلیل نتایج

تحلیل موضوعی احساسات:

- اپلیکیشن موبایل: 65% نظرات منفی (بیشتر به دلیل باگ‌ها و قطع اتصال)
- خدمات مشتریان: 45% مثبت، 35% خنثی، 20% منفی
- بانکداری آنلاین: 70% احساس مثبت به‌خاطر سهولت استفاده
- دستگاه‌های ATM: احساسات ترکیبی بسته به منطقه

تحلیل خطا:

- مشکل در شناسایی طعنه و کنایه
- جملات دو‌زبانه (فارسی/انگلیسی)
- اصطلاحات خاص بانکی و لهجه‌های محلی

ابزارها و تکنولوژی‌ها

- زبان برنامه‌نویسی: Python 3.8

- کتابخانه‌ها:

- NLP: hazm, transformers, scikit-learn, torch

- Visualization: matplotlib, seaborn, plotly

- Web scraping: requests, beautifulsoup4, selenium

نتیجه‌گیری

با استفاده از روش‌های پیشرفته در پردازش متن فارسی و مدل‌های یادگیری عمیق، سیستم طراحی شده توانست به دقت بالایی در تحلیل احساس کاربران نسبت به خدمات بانکی دست پیدا کند. مدل Ensemble نهایی عملکردی بسیار مطلوب داشت و تحلیل موضوعی نیز نشان‌دهنده نیاز به بهبود در حوزه‌های خاصی از خدمات بانکی است.

پیشنهاد برای آینده

- استفاده از مدل‌های جدیدتر مانند RoBERTa یا GPT برای زبان فارسی
- تحلیل زمانی برای بررسی تغییر احساسات در طول زمان
- افزودن نظرات از شبکه‌های اجتماعی دیگر (مثل توییتر و اینستاگرام)
- ساخت داشبورد تحلیل احساسات برای استفاده مدیران بانکی