



دانشگاه شهید بهشتی  
دانشکده علوم پایه

پیشنهادیه پروژه درس نظریه یادگیری ماشین

تحلیل احساسات توییت های شرکت های هواپیمایی

## Sentiment Analysis of Airline Twitter data

دانشجو: زهرا اروجی

استاد: دکتر حسین حاجی ابوالحسن

مقطع: کارشناسی ارشد

گرایش: علوم داده ها

تیرماه ۱۳۹۹

صفحه

فهرست عناوین

۱	مقدمه و طرح مسأله.....	۳
۲	تاریخچه.....	۳
۳	شرح داده.....	۳
۴	شرح روش.....	۴
۵	منابع و مراجع.....	۵

## ۱ مقدمه و طرح مسأله

در دنیای امروز، اینترنت حجم وسیعی از اطلاعات را در بر گرفته است. افراد اینترنت را به عنوان منبع مهمی درک کرده اند که در آن تعداد زیادی از نظرات و تجربیات به آسانی در دسترس هستند. ارزیابی های مردم به طور قابل توجهی بر باورها، برداشتها و به ویژه تصمیمات خرید آنها تأثیر می گذارد. امروزه، جریان و گردش اطلاعات به تدریج، به تجمع آنلاین تجربیات، بینشها و دیدگاهها تبدیل شده است. افزایش شدید اطلاعات آنلاین فرصت قابل توجهی را برای شرکتها ایجاد می کند تا بهتر درک کنند که مشتریان درباره یک محصول، موضوع یا نهاد دیگر چه می گویند.

در سالهای اخیر، تجزیه و تحلیل احساسات تویتر برای تجزیه و تحلیل خودکار رضایت مشتری از خدمات آنلاین بسیار رایج شده است. بازخورد مشتری در مورد خدمات آنها، برای شرکتها مخصوصاً شرکت های هواپیمایی بسیار ضروری است.

در این پروژه، داده های تویتر چند شرکت هواپیمایی مورد تجزیه و تحلیل قرار گرفته است که عواملی چون بهترین و بدترین خطوط هواپیمایی، بیشترین دلایل نارضایتی مشتریان، نظرات مثبت، منفی و خنثی آنها درباره ی خدمات شرکت های هواپیمایی، در محیط google colab با استفاده از الگوریتم های یادگیری ماشین، پیش پردازش، بصری سازی، ارزیابی و توسط شبکه عصبی مدل و پیش بینی می شود.

## ۲ تاریخچه

با رشد سریع شبکه های اجتماعی و بحث و گفتگوی آنلاین، وب غنی از داده های متن آزاد تولید شده توسط کاربر است، جایی که کاربران می توانند نگرش های مختلفی را نسبت به محصولات ابراز کنند. که این موضوع باعث شده است محققان به سمت تحلیل احساسات جذب شوند.

برای تعیین اینکه یک متن یا یک جمله بیانگر احساسات مثبت یا منفی باشد، معمولاً از دو رویکرد اصلی استفاده می شود: رویکرد مبتنی بر واژگان و رویکرد مبتنی بر یادگیری ماشین. رویکرد مبتنی بر واژگان شامل محاسبه جهت گیری برای یک متن از جهت معنای کلمات یا عبارات موجود در متن است. رویکرد مبتنی بر یادگیری ماشین شامل ایجاد classifiers از نمونه های دارای برچسب متن ها یا

جملات است. اغلب این تکنیک‌ها مانند Maximum Entropy, Naive Bayes و Support Vector Machines در زمینه supervised learning هستند.

بطور مثال Sreenivasan و همکاران روی تویت‌های ۳ شرکت هواپیمایی مطالعه کردند. آنها از تویت‌ها به عنوان منبع داده برای تحلیل ارتباطات مصرف کنندگان در مورد خدمات هواپیمایی استفاده کردند. همینطور Breen و همکاران طبقه‌بندی احساسات تویت‌ها را با استفاده از واژه‌نامه احساساتی نمایش دادند و پیشنهاد دادند تویت‌های دارای زمان واقعی از Twitter API به جای پرسش‌های حاوی نمایش نام شرکت‌های هواپیمایی بازیابی شوند. در این روش آنها به دقت ۸۶/۴٪ رسیدند. و همچنین Adeborna و همکاران در رویکرد Naive Bayesian، دو روش SVM و Entropy را مقایسه کرده اند.

### ۳ شرح داده

مجموعه داده مورد نظر ما که از سایت آموزشی kaggle دریافت شده است به فرمت csv می‌باشد که شامل ۱۴۶۴۰ تویت از ۷۷۰۰ کاربر و مجموعاً ۱۵ ویژگی است. داده‌های تویت‌ها از فوریه ۲۰۱۵ حذف شد و از مشتریان خواسته شد ابتدا تویت‌های مثبت، منفی و خنثی را ثبت کنند و به دنبال آن دسته بندی دلایل منفی (مانند "پرواز دیررس" یا "سرویس دهی بد") را اشتراک گذاری کنند.

### ۴ شرح روش

در ابتدا کتابخانه‌های مورد نیاز و اساسی همچون scikit learn و keras و همچنین مجموعه داده را فراخوانی می‌کنیم. سپس به پردازش و پاک سازی داده و استخراج ویژگی‌های کاربردی یا استفاده از تحلیل نمودارها و درک مجموعه داده‌ها می‌پردازیم. سپس با استفاده از روش امتیازدهی به نظرات مثبت و منفی، مجموعه داده‌ها را برای آموزش آماده کرده و در نهایت الگوریتم منتخب یادگیری ماشین را روی مجموعه داده‌های حاصل به کار می‌گیریم. در نهایت به ارزیابی خروجی مدل می‌پردازیم.

روش به این گونه است که شبکه عصبی را روی داده‌های اجرا می‌کنیم و با بررسی مدل به این نتیجه می‌رسیم که مدل ما به دقت ۹۴٪ برای تویت‌های منفی و دقت ۸۲٪ برای تویت‌های مثبت می‌رسد که این دقت نسبت به مدل‌های دیگر یادگیری ماشین که توسط اینجانب بررسی شده دارای میزان خوب و قابل قبول تری است.

## منابع و مراجع

- [١] Haji H. BINALI, Chen WU, Vidyasagar POTDARA, “New Significant Area: Emotion Detection in E-learning Using Opinion Mining Techniques”, 3rd IEEE International Conference on Digital Ecosystems and Technologies, pp. 259-264, 2009
- [٢] Bing Liu. Sentiment analysis and subjectivity. In Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca, 2010.
- [٣] Hsu CW, L.C., A Simple Decomposition Method for Support Vector Machines. Machine Learning, 46, 291–314. URL <http://www.csie.ntu.edu.tw/~cjlin/papers/decomp.ps.gz>, 2002
- [٤] Nan, L.a.D., D., Using text mining and sentiment analysis for online forums hotspot detection and forecast. Decision Support Systems archive. Volume 48 Issue 2, January, 2010. Pages 354-368. Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, 2010.
- [٥] Sreenivasan, Nirupama Dharmavaram, Chei Sian Lee, and Dion Hoe-Lian Goh. "Tweeting the friendly skies: Investigating information exchange among Twitter users about airlines." Program: electroni.
- [٦] Breen, Jeffrey Oliver. "Mining twitters for airline consumer sentiment." Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications , pp. 133, 2012
- [٧] Adeborna, Esi, and Keng Siau. "An Approach to Sentiment Analysis–The Case of Airline Quality Rating." PACIS 2014 Proceedings, pp.363.