کنگره ملی توسعه پژوهش های نوین در مهندسی برق و کامپیوتر
موسسه آموزش عالی وحدت، تربت جام، ایران
۸و ۱۹اسفند ۹۷
2nd National Congress of Innovative Research Developments in Electrical and Computer Engineering

به‌کارگیری روش متن‌کاوی به‌منظور دسته‌بندی سلسله مراتبی متون با رویکرد یادگیری عمیق درداده های حجیم

بهناز اسلامی[۱]، مهدی حبیب زاده مطلق[۲*]، مجید فولادیان[۳] و زهرا رضایی[۴]
[۱] گروه مهندسی کامپیوتر، دانشکده برق و کامپیوتر، واحد علوم تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران، behnazeslami30@gmail.com
[۲] استادیار گروه کامپیوتر، دانشگاه ایوانکی، سمنان، ایران، me_habi@encs.concordia.ca (نویسنده مسئول)
[۳] استادیار گروه برق و الکترونیک، دانشگاه آزاد اسلامی واحد ساوه، ساوه، ایران، majidfouladian@gmail.com
[۴] گروه مهندسی کامپیوتر، دانشکده برق و کامپیوتر، دانشگاه کاشان، اصفهان، ایران، z.rezaei2010@gmail.com

چکیده - همواره تجزیه‌وتحلیل متن یک روش برای استخراج دانش از متن است که جهت بازیابی دانش نهان در متون، محدودیت حافظه و زمان در پردازش داده‌های بزرگ به دلیل منابع داده توزیع‌شده در وب، موتورهای جستجو و شبکه‌های اجتماعی حائز اهمیت می‌باشد. با بهبود داده‌ها در اینترنت و شبکه‌های اجتماعی، روش‌های قوی‌تر موردنیاز است که می‌تواند اطلاعات را از نزدیک و بلافاصله دسته‌بندی کند. از طریق انتزاع درداده های متنی، یادگیری عمیق می‌تواند با این چالش‌ها مقابله کند. در این مقاله روش یادگیری عمیق جهت دسته‌بندی به‌صورت سلسله مراتبی موردبررسی قرار می‌گیرد و نتایج نشان می دهند که این روش می تواند متون شبکه‌های اجتماعی و وب‌سایت‌ها را با دقت ۹۸,۸۱٪ دسته‌بندی کند. این مقاله همچنین نشان می‌دهد که این شبکه پیچیده می‌تواند در سطح کلمه بهتر کار کند و نیازی به دانستن ساختار نحوی و معنایی زبان ندارد. نوآوری روش پیشنهادی، اضافه کردن یک سطح به ساختار سلسله مراتبی برای تشخیص عمومی و تشخیص دقیق‌تر کلاس است.

کلیدواژه- پردازش زبان طبیعی، داده‌های حجیم، دسته‌بندی، متن‌کاوی، یادگیری عمیق

۱- مقدمه

دسته‌بندی متن به‌عنوان یکی از روش‌های اصلی در پردازش طبیعی فرآیندهای زبان در نظر گرفته‌شده است و هدف این است که یک متن خاص را برچسب‌گذاری کند. پردازش دستی و دسته‌بندی داده‌ها برای کاربران انسان غیرعملی است، زیرا اطلاعات زیادی وجود دارد [۱].

به‌علاوه تهیه مجموعه از متن‌هایی که برای آموزش چنین مدلی استفاده می‌شود، به‌عنوان یک چالش در نظر گرفته می‌شود. در این میان، کارشناسان موضوعی قادر هستند تا اسناد را تشخیص داده و عمل برچسب‌گذاری را انجام داده و مجموعه آموزشی را ایجاد کنند. انجام این فرآیند، زمان و انرژی زیادی را صرف می‌کند. علاوه بر این، تعداد اسناد متنی موردنیاز برای ساخت یک مدل ناشناخته است، درحالی‌که ایده کلی وجود اسناد متنی بیشتر جهت ارائه مدل بهتر می‌باشد [۲]. نکته دیگر که باید اشاره کرد این است که اکثر ابزار تجزیه‌وتحلیل متن از پردازش خطی و خطی پشتیبانی نمی‌کند.

علاقه به پردازش داده‌ها در دنیای امروزی توجه زیادی به مدل‌های داده بزرگ مانند Apache Spark [۳]،Apache Storm [۴]و Apache Flink [۵] را به خود جلب کرده است. ابزارهای دیگر مانند Azure ML [۶] و SAMOA [۷] نیز ازجمله ابزارهای بسیار کارآمد هستند که Azure ML تنها برای خدمات ابری ارائه می‌شود؛ که کاربران را مسدود می‌کند و از سوی دیگر، SAMOA فقط از داده‌های بلادرنگ پشتیبانی می‌کند و نمی‌تواند برای همه برنامه‌ها اعمال شود.

به طور کلی متن‌کاوی شامل بخش‌های مختلفی می‌شود که برخی از آن‌ها عبارتند از : مدل‌سازی موضوع [۸]، دسته‌بندی احساسات [۹] و تشخیص هرزنامه [۱۰]. در گذشته با استفاده از روش‌های سنتی دسته‌بندی متن، ابتدا متن‌ها به‌صورت نمودارهای n-gram و ویژگی‌های لغوی نمایش داده می‌شدند،

سپس یک مدل خطی یا روش هسته بر روی این نمودارها اعمال می‌گردید[۸]. اما اخیراً ارائه متنوع متون با استفاده از روش‌های یادگیری عمیق همانند شبکه‌های عصبی کانولوشن [۱۱] یا بر اساس شبکه‌های عصبی مکرر که خودشان بر اساس حافظه کوتاه‌مدت [۱۲] ارائه می‌شوند ارائه‌شده است. دسته‌بندی و پیش‌بینی دو عملیات برای تجزیه‌وتحلیل داده‌ها و مشتق مدل‌ها باهدف توصیف دسته‌های اصلی داده‌ها، درک آن‌ها و پیش‌بینی رفتار آینده آن‌هاست. به‌طورکلی، دسته‌بندی داده‌ها یک فرآیند دومرحله‌ای است. مرحله اول، ساخت مدل است و مرحله دوم، استفاده از مدل و پیش‌بینی از طریق داده‌های قبلی است. دسته‌بندی اسناد به این معنی است که دسته‌بندی این اسناد و متون را به گروه‌هایی با ویژگی‌های مشترک، به‌طوری‌که بعداً می‌توان از این اسناد بر اساس این نقطه مشترک استفاده کرد.

ازآنجاکه اسناد ممکن است بر اساس محتوا یا بر اساس ویژگی دیگری مانند تاریخ اسناد دسته‌بندی شوند. در دسته‌بندی محتوا اسناد، ابتدا برخی اسناد با موضوع پیش تعیین‌شده به‌عنوان نمونه برای یادگیری به دستگاه داده می‌شوند. سپس، بر اساس کلمات در هر سند، شبکه یاد می‌گیرد و با توجه به‌احتمال کلمات در هر سند، دستگاه موضوع سایر اسناد را پیش‌بینی می‌کند. سپس، با استفاده از یادگیری عمیق دسته‌بندی به‌صورت سلسله مراتبی انجام‌شده و نتایج ارزیابی خواهد شد. نکته مهم در این مقاله معرفی این است که دسته‌بندی فارغ از در نظر گرفتن چالش‌های داده‌های بزرگ و گفتار محاوره‌ای انجام می‌پذیرد. مراحل کلی این مقاله به شرح زیر است:

- جمع‌آوری داده‌ها
- پیش‌پردازش
- دسته‌بندی داده‌های حجیم به روش سلسله مراتبی

۲- مطالعات قبلی

در [۱۲] نشان داده‌شده است که آموزش شبکه عصبی کانولوشن بر بردارهای کلمات که قبلاً ایجادشده است، برای دسته‌بندی جمله بسیار مؤثر است. علاوه بر این، یک شبکه عصبی کانولوشن تک لایه نیز عملکرد بهتری در مقایسه با الگوریتم‌های بیزین و ماشین بردار پشتیبان دارد. علاوه بر این، در این مقاله، شبکه‌های عصبی کانولوشن برای استفاده از استخراج ویژگی از متن استفاده‌شده است. در [۱۳] برخی از ویژگی‌های داده‌های متنی از طریق ترسیم شبکه عصبی کانولوشن ساخته‌شده‌اند و این ویژگی‌ها با استفاده از یک بردار ۳۶۰ درجه‌ای که شامل بردار کلمه و نوع ساختار کلمه می‌باشد نشان داده‌شده‌اند. در این مقاله، خروجی انتهایی یک‌لایه‌ی کاملاً متصل در شبکه عصبی کانولوشن همراه با سایر طبقه‌بندی‌ها استفاده‌شده است. همچنین در این مقاله مشخص شد که ویژگی‌های ایجادشده توسط شبکه، دسته‌بندی بهتر داده‌های آموزشی را از طریق سیستم‌های مختلف یادگیری ارائه می‌دهد. علاوه بر این مشخص گردید، ویژگی‌های ایجادشده بهتر است جهت دسته‌بندی به لایه آخر شبکه عصبی متکی باشند. در [۱۴] نشان داده‌شده که دقت دسته‌بندی متون با استفاده از تکنیک‌های پیشرفته پردازش زبان طبیعی به‌طور قابل‌توجهی بهبود بخشیده شده است. در این مطالعه، گروهی از نظرات پزشکی در مورد یک دسته از داروها برای بررسی اثرات جانبی داروها از طریق ویژگی‌های مفید آن‌ها موردبررسی قرار گرفت. نتایج نشان داد که انتخاب دقیق مشخصات و ویژگی‌ها می‌تواند به‌طور خودکار نتایج نهایی برخی از روش‌های دسته‌بندی کلاسیک ازجمله بیزین و ماشین بردار پشتیبان را بهبود بخشد.

۳- روش پیشنهادی

روش پیشنهادی شامل دو مرحله است : جمع آموری داده‌ها جهت ساخت مدل و پیش‌بینی بلادرنگ متون. در این بخش، این مراحل شرح داده می‌شوند.

۳-۱- جمع‌آوری داده‌ها

با استفاده از خزش گر Scrapy متون مختلف از وب‌سایت‌های خبری CNN و رویترز در ۱۱ دسته در بازه زمانی سال ۲۰۱۶ تا سال ۲۰۱۷ بر اساس موضوعات مختلف جمع‌آوری شدند.

۳-۲- پیش‌پردازش متون

- نشانه‌گذاری: تبدیل یک سند به یک گروه از کلمات ($W_0, W_1, ..., W_n$) که در آن n فرکانس کلمات منحصربه‌فرد در یک داکیومنت متنی است.
- تبدیل حروف بزرگ به کوچک: تمام کلمات را از حروف بزرگ به حروف کوچک تبدیل می‌کند. بر این اساس، کلمات مانند "Runner" یا "runner" به معنای دونده، هر دو به‌عنوان کلمه "runner" شناسایی می‌شوند
- حذف علائم زائد : به معنای حذف علائمی همانند: ، ; "" "! @ # $ ٪ ^ , * (<>؟ و موارد دیگر.

جدول۱ : اطلاعات دسته‌ها به تفکیک موضوع

زبان	دسته	شماره برچسب
انگلیسی	هنر	۰
انگلیسی	ورزشی	۱
انگلیسی	پزشکی	۲
انگلیسی	مذهبی	۳
انگلیسی	خبری	۴
انگلیسی	سیاسی	۵
انگلیسی	فرهنگی	۶
انگلیسی	اقتصادی	۷
انگلیسی	اجتماعی	۸
انگلیسی	تاریخی	۹
انگلیسی	سایر	۱۰

جدول۲ : نمونه‌ای از پیش‌پردازش

	اعمال پیش‌پردازش
متن	“Christopher Columbus”, the earliest European visitor, made his first land Farsill in the New World on the Bahamian island of San Salvador in 1492!!!
تبدیل حروف بزرگ به کوچک	“christopher columbus”, the earliest european visitor, made his first land farsill in the new world on the bahamian island of san salvador in 1492!!!
حذف علائم زائد	christopher Columbus the earliest european visitor made his first landFarsill in the new world on the bahamian island of san salvador in 1492

۳-۳- معماری روش پیشنهادی

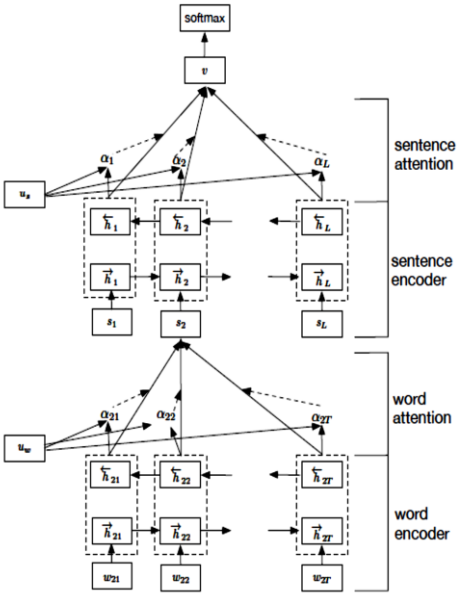
در این مقاله، داده‌های ورودی به‌صورت جداگانه به معماری پیشنهادی افزوده می‌شوند. معماری پیشنهادی بر اساس ساختار شبکه‌های عصبی است و ورودی آن لزوماً فرمت متن است و بر اساس مکانیزم داخلی آن متون را به بردارهای عددی تغییر می‌دهد و فرایند خود را دنبال می‌کند. به‌منظور کاهش اشتباهات از روش سلسله مراتبی مبتنی بر یادگیری عمیق در شبکه عصبی استفاده می‌شود. در حقیقت هدف معرفی روشی است که در آن خطای دسته‌بندی به حداقل برسد. اگرچه روش‌های دسته‌بندی متون بر مبتنی بر شبکه‌های عصبی دارای راندمان بالایی هستند [۱۷-۱۵]. لذا فرضیه تحقیق حاضر این است که اگر از دانش ساخت متون در ساختار مدل تولیدشده استفاده کنیم، متن بهتری به دست می‌آید. ایده ما در ارائه این مدل این است که تمام اجزای متن در پاسخ به جستجو یک وابستگی یکسان ندارند. علاوه بر این، شناسایی بخش‌های مختلف روابط بین کلمه هم وابسته است.

روش پیشنهادی یک ساختار عصبی جدید یا شبکه سلسله مراتبی است که برای انجام دو ویژگی اصلی در مورد ساختار متن طراحی‌شده است. نخست آن‌که چون متون ساختار سلسله مراتبی دارند (ازجمله کلمات عبارت و عبارات یک متن و در مقیاس بزرگ متن از مجموعه‌ای از اسناد)، به همین دلیل نمایش متون ابتدا با عبارات آغاز می‌شود و از کنار هم قرار گرفتن عبارات متن نهایی حاصل می‌شود. همچنین، مشاهده‌شده است که کلمات و عبارات متن متفاوت متون و بارهای مختلف هستند. علاوه بر این، اهمیت کلمات و عبارات عمیقاً وابسته به زمینه‌ای است که از آن‌ها استفاده می‌شود. به‌طوری‌که همان کلمه ممکن است به‌طور کاملاً متفاوت در یک زمینه دیگر بامعنای دیگری به کار برود. به همین دلیل به‌منظور در نظر گرفتن این حساسیت، مدل پیشنهادی در دو سطح مکانیسم توجه قرارگرفته است: یکی در سطح کلمه و دیگری در سطح عبارت.

ساختار درروش پیشنهادی سلسله مراتبی است و مکانیزم آن در ابتدا در سطح کلمه و سپس در سطح جمله و درنهایت در سطح سند است. درواقع، این به این دلیل است که اهمیت کلمات و جملات وابسته به زمینه‌ای است که در آن‌ها استفاده

می‌شود، به‌طوری‌که یک کلمه یا جمله ممکن است یک مقدار کاملاً متفاوت در یک زمینه متفاوت داشته باشد. بنابراین، سه سطح مکانیسم در نظر گرفته‌شده است، و این به ما امکان می‌دهد بخش‌های مهم‌تر متن را از قسمت‌های مهم کمتر درحالی‌که ساخت یک متن خاص ارائه می‌دهد، تمایز بخشد.

ساختار کلی شبکه سلسله مراتبی مورداستفاده (HAN) در نمودار زیر ارائه‌شده است که شامل اجزای مختلف است: رمزگذار و توجه در سطح کلمه و رمزگذار و توجه در سطح جمله و درنهایت اضافه کردن رمزگذار و توجه به‌کل سند.



شکل ۱: معماری روش HAN [۱۸]

از طریق روش سلسله مراتبی استفاده از بردارهای کلمات (Word2vec) بر اساس عملکرد مکانیک رمزگذار و توجه در سطوح کلمه و جمله به‌صورت متقابل (یک‌بار از انتها به ابتدای و یک‌بار از ابتدا تا انتها)، شباهت‌ها شناسایی می‌شوند و در زمان آموزش، دقت آن در تشخیص افزایش می‌یابد. به‌طورکلی می‌توان گفت که آموزش شبکه HAN، به‌شدت به بردارهای کلمات (Word2Vec) بستگی دارد و این بردارها هرقدر غنی‌تر باشند آموزش بهتر انجام می‌گیرد. درواقع، آنچه در این روش پیشنهادی مهم است، جایگزینی عملکرد پیش‌فرض آن از tanh به sigmoid است که بر دقت یا خطای خروجی تأثیر می‌گذارد[۱۸].

مکانیزم توجه در شبکه‌های عصبی یک کپی از مکانیزم توجه بصری در انسان است، به این معنی که توجه بصری انسان می‌تواند در یک نقطه خاص از یک تصویر با "وضوح‌بالا" متمرکز شود درحالی‌که در اطراف آن با "وضوح پایین" درک می‌شود و سپس / او می‌تواند نقطه کانونی را در طول زمان تنظیم کند. مکانیزم توجه در NLP، مدل را باقابلیت یادگیری آنچه مطابق با متن ورودی و آنچه تاکنون تولید کرده است، یاد می‌گیرد.

بر این اساس، از طریق مکانیزم توجه، ما کلمات را که در درک متن بسیار مهم هستند، شناسایی می‌کنیم، سپس واژه‌هایی را که ازلحاظ معنایی غنی هستند را به شکل یک بردار تبدیل می‌کنیم.

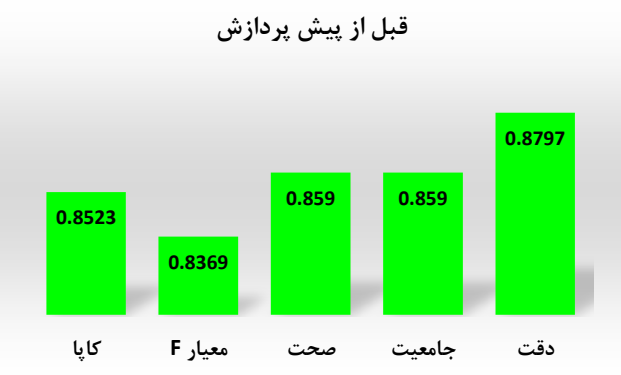
به‌منظور تشخیص جملات که سرنخ برای دسته‌بندی یک متن هستند، اهمیت هر جمله از طریق یک بردار و با استفاده از یک بردار مشخص می‌شود. سپس، با توجه به جملات در اسناد متنی، احتمال قوی‌تر از دسته‌بندی متن چند متن می‌تواند به دست آید، که چقدر از متن یک موضوع خاص را ارائه می‌دهد و چقدر از آن موضوع دیگری را نشان می‌دهد. هدف از انجام این کار یافتن متون مبتنی بر وب است که از کلاس‌های مختلف تشکیل‌شده و گاهی درون یک متن، مثلاً متن اقتصادی، مسائل سیاسی مطرح می‌شود. که در این حالت سطح وابستگی متن به هر کلاس مهم است.

۴- نتایج

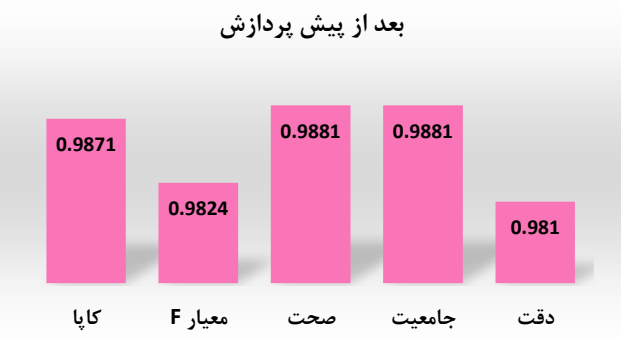
در این مقاله مجموع کل داده‌ها در بخش آموزش شامل ۱۹۶۰۱۰ سند و ۲۲۰۰۰ سند برای بخش آزمون بود. تمامی داده‌های آموزش و آزمون به‌صورت دوره‌ای به‌صورت جداگانه به شبکه وارد می شدند که سیستم فرآیند دسته‌بندی را در کمتر از ۰,۰۱۴ انجام داد. مرحله آزمایش روی GEFORCE GTX 1080 انجام گردید، که یک‌میلیون سند متنی در هر ثانیه دسته‌بندی شدند. این پژوهش بر اساس زبان پایتون با استفاده از ابزارهای ویژه مانند کتابخانه‌های جستجوی متن مانند NLTK، Genism و TensorFlow صورت گرفته است.

اثر پیش‌پردازش داده‌ها در تصاویر زیر نشان داده‌شده است، همان‌طور که مشخص است، با توجه به این واقعیت که داده‌ها از تمام وب‌سایت‌های گرفته‌شده‌اند، بنابراین آن‌ها شامل برخی بخش‌های غیر مرتبط هستند. ازآنجاکه تنها بخش‌های

کارآمد در هر خبر تنها همان متن خبرمی باشد بخش‌های دیگر تأثیر منفی بر خروجی دارند به همین دلیل در مرحله پیش‌پردازش این موارد زائد شناسایی‌شده و حذف می‌شوند.



شکل ۲: نمودار میله‌ای معیارهای ارزیابی قبل از پیش‌پردازش



شکل ۳: نمودار میله‌ای معیارهای ارزیابی بعد از پیش‌پردازش

با توجه به زمان صرف شده برای دسته‌بندی متن، با توجه به این‌که یک صفحه استاندارد دارای ۱۵۰۰ کلمه است، سند متون مربوط به آزمون حاوی ۲۰۵۶ صفحه بود که درنهایت به‌منظور آزمودن سرعت تجزیه‌وتحلیل، سرعت دسته‌بندی کل متون مورد آزمایش ۲۰,۱۴۸ ثانیه بود که به‌طور تقریبی می‌توان گفت که هر صفحه در ۰,۰۱۴۱۷۷۰۴۳ ثانیه برای تجزیه‌وتحلیل مورد دسته‌بندی قرار می‌گرفت.

۵- نتیجه‌گیری و بحث

در این مقاله از یک مدل یادگیری عمیق برای دسته‌بندی سلسله مراتبی گسترده استفاده شد. در ابتدا کلمات موجود در سطح جملات متون موجود به‌صورت (BOW) استخراج شدند، سپس با استفاده از قدرت بالای ترکیب معنایی برای دسته‌بندی استفاده شد. در مقایسه با روش‌های سنتی و مدرن دسته‌بندی متن، رویکرد کنونی می‌تواند نتایج دسته‌بندی اسناد به زبان انگلیسی را بهبود بخشد. در این مقاله شبکه‌های توجه سلسله مراتبی معرفی گردید.. این مدل درنهایت بردارهای سند را از طریق ایجاد یک بردار متشکل از کلمات در بردارهای جمله می‌سازد که نتیجه همگرایی بردارهای کلمات مهم بوده و درنهایت، بردارهای سند تشکیل می‌شوند. مدل پیشنهادی روش‌های قبلی را به‌طور قابل‌توجهی بهبود می‌بخشد. ارائه لایه‌های به‌دست‌آمده توجه نشان می‌دهد که مدل عملکرد مناسبی و قابل‌توجهی در انتخاب کلمات و جملات داشته است. علاوه بر این، در متن‌هایی که احتمال آن‌ها کمتر از آستانه تشخیص است، استفاده از حافظه کوتاه‌مدت در RNN می‌تواند به مدل کمک کند تا مدل‌ها را به‌طور مؤثرتر تشخیص دهد. نقطه نهایی این است که در این روش، تعداد پارامترهایی که در مرحله آموزش به‌روز می‌شوند کمتر گردیده و این خود باعثمی شود مدت‌زمان فاز آموزش کاهش‌یافته و با سرعت بالاتری صورت پذیرد. ازاین‌رو این روش می‌تواند جهت دسته‌بندی متون در مباحث داده‌های حجیم راه‌حل مناسبی باشد.

مراجع

[1] Stahl F, Gaber MM, Adedoyin-Olowe M. “A Survey of Data Mining Techniques for Social Media Analysis”. *Journal of Data Mining & Digital Humanities*.2014.
[2] Kwon O, Sim JM. “Effects of data set features on the performances of classification algorithms”. *Expert Systems with Applications*. Vol. 40, No. 5, pp. 1847-1857, 2013.
[3] Meng X, Bradley J, Yavuz B, Sparks E, Venkataraman S, Liu D, Freeman J, Tsai DB, Amde M, Owen S, Xin D. “Mllib: Machine learning in apache spark”. *The Journal of Machine Learning Research*. Vol. 17, No. 1, pp. 1235-1241, 2016.
[4] Allen ST, Jankowski M, Pathirana P. “Storm Applied: Strategies for real-time event processing”. *Manning Publications Co*.; 2015.
[5] Carbone P, Katsifodimos A, Ewen S, Markl V, Haridi S, Tzoumas K. “Apache flink: Stream and batch processing in a single engine”. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*. Vol. 36, No. 4, 2015.
[6] Barga R, Fontama V, Tok WH, Cabrera-Cordon L. “Predictive analytics with Microsoft Azure machine learning”. *Apress*, 2015.
[7] De Francisci Morales G. “SAMOA: A platform for mining big data streams”. In: *Proceedings of the 22nd International Conference on World Wide Web*. pp. 777-778, 2013.
[8] Wang, S., & Manning, C. D. “Baselines and bigrams: Simple, good sentiment and topic classification”, In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistic*. Vol. 2, pp. 90-94, 2012.

[9] Pang and Lillian Lee, "Opinion mining and sentiment analysis", *Foundations and trends in information retrieval*. Vol. 2, No. 1-2, pp. 1–135, 2008.
[10] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E, "A Bayesian approach to filtering junk e-mail. In Learning for Text Categorization". *Papers from the 1998 workshop*, Vol. 62, pp. 98-105, 1998.
[11] Blunsom, P., Grefenstette, E., & Kalchbrenner, N., "A convolutional neural network for modelling sentences", In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014.
[12] Hochreiter, S., & Schmidhuber, J. "Long short-term memory", *Neural computation*. Vol. 9, pp. 1735-1780, 1997.
[13] Y. Kim, "Convolutional Neural Networks for Sentence Classification", *Empirical Methods in Natural Language Processing (EMNLP)*. Doha, 2014.
[14] S. Poria, E. Cambria and A. Gelbukh, "Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-Level Multimodal Sentiment Analysis". *Empirical Methods in Natural Language Processing*. Lisbon, 2015.
[15] A. Sarker and G. Gonzalez, "Portable automatic text classification for adverse drug reaction detection", *Journal of Biomedical Informatics*. Vol. 53, pp. 196-207, 2015.
[16] Zhang X, Zhao J, LeCun Y. "Character-level convolutional networks for text classification". In: *Advances in neural information processing systems*. pp. 649-657, 2015.
[17] Johnson R, Zhang T. "Effective use of word order for text categorization with convolutional neural networks". ArXiv preprint arXiv:1412.1058. 2014.
[18] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola and E. Hovy, "Hierarchical Attention Networks for Document Classification", in *NAACL-HLT*, San Diego, 2016.