تشخیص موضوع متون خبری فارسی با روش دمپسترشفر

علی جبار رشیدی [۱]، ملک نظرانداز [۲]

۱- دانشیار- دانشگاه صنعتی مالک اشتر- مجتمع دانشگاهی برق و کامپیوتر aiorashid@yahoo.com

۲-دانشجوی کارشناسی ارشد رایانش امن - دانشگاه صنعتی مالک اشتر nazarandaz.m@gmail.com

چکیده

با رشد سریع و روز افزون اطلاعات، رده‌بندی مستندات یکی از ابزارهای کلیدی برای سازماندهی و مدیریت داده‌های متنی به شمار می‌آید که در کاربردهایی مانند تقسیم‌بندی اخبار، نامه‌های الکترونیکی و اطلاعات آنلاین مورد استفاده قرار می‌گیرد. در واقع رده‌بندی موضوعی متون، تعیین موضوع یک متن می‌باشد. نظرات حاوی اطلاعات ارزشمندی هستند که می توان با نظرکاوی آن ها به دانش ارزشمندی در ارتباط با یک موضوع خاص دست یافت چراکه تمرکز نظرکاوی برروی تجزیه و تحلیل یک متن برای درک نظرات بیان شده است. در این تحقیق نیز از نظر کاوی مبتنی بر روش دمپستر شفر استفاده شد تا اطلاعات ارزشمندی از متن استخراج شود.

هدف کلی مقاله بهبود روش های نظر کاوی با روش دمپستر شفر است. با کمک انواع طبقه بندی یادگیری ماشین یک سامانه نظر کاوی میتواند به طور موثر اطلاعات مرتبط با نظرات را به زبان طبیعی درک نماید. با وجود کارهای خوب صورت گرفته در زمینه متون فارسی، هنوز برخی از چالش‌ها به صورت حل نشده باقی مانده‌اند. از جمله این چالش‌ها استفاده از ادغام تصمیم از متون برای طبقه بندی آن‌ها می‌باشد. بسیاری از منابع از روش های مستقل برای نظر کاوی استفاده نموده اند که کمتر به ترکیب روش ها توجه می کند. با اجرای طرح پیشنهادی و انجام آزمایشات متعدد، کارایی و اثر بخشی مدل پیشنهادی در سطح جمله به دقت ۹۲٪ رسیده است.

کلمات کلیدی: تشخیص موضوع ، نظرکاوی ، روش دمپستر شفر ، یادگیری ماشین ، ادغام تصمیم ، شبکه های سایبر اجتماعی

۱. مقدمه

با گسترش سریع متون الکترونیکی که همراه با ساختارها و زبان‌های متفاوتی بودند، توجه بسیاری از دانشمندان و محققان علوم کامپیوتر به استفاده از روش‌ها و تکنیک‌های بهینه و سریع برای دسته‌بندی متون الکترونیکی جلب شد و هم اکنون نیز تحقیق در این زمینه در راستای افزایش سرعت و دقت روش‌ها همچنان ادامه دارد [۱]. انواع مختلفی از اطلاعات بارگذاری و به اشتراک گذاری در شبکه های سایبر اجتماعی در قالب متن ، فیلم ، عکس و صدا وجود دارد [۲]. رسانه های اجتماعی سرشار از داده های خام و پردازش نشده و بهبود فناوری ، به ویژه در یادگیری ماشین و هوش مصنوعی است ، اجازه می دهد تا داده ها پردازش شوند و آن را به یک داده مفید تبدیل کنند کاربران می توانند به سرعت و صریح احساسات خود را بدون نیاز به تایپ پیام های توصیفی طولانی انتقال دهند .[۳]

رده‌بندی یا طبقه‌بندی متون[۱] شامل اختصاص یک برچسب به یک سند می‌باشد که می‌تواند به چندین منظور صورت گیرد. نوعی از این رده‌بندی که در نظرکاوی کاربرد دارد، تحلیل احساسات بوده و به مجموعه‌ای از متون سه برچسب را اختصاص می‌دهد که شامل برچسب‌های منفی، مثبت و خنثی می‌باشد. البته این نوع رده‌بندی به صورت دو کلاسه (مثبت و منفی) نیز قابل انجام است.این سه یا دو دسته در برنامه های مختلف مفید هستند [۴] اما به هر حال جزئیات دقیق مربوط به احساسات بیان شده در پیام را ارائه نمی دهند. به عنوان مثال ، نظرکاوی نمی تواند احساساتی مانند شادی ، ترس ، تعجب ، احساس گناه و غیره را درک کند ، که برای درک احساس واقعی ابراز شده توسط کاربر ضروری است. برای تفسیر موثرتر پیام ، نه تنها شناسایی محتویات و متن پیام ضروری است ، بلکه تشخیص احساسات بیان شده در پیام نیز ضروری است. [۵] اما نوع دیگر رده‌بندی، رده‌بندی موضوعی متن می‌باشد. در این نوع رده‌بندی، هدف اختصاص یک برچسب به متن با توجه به موضوع آن بوده که شامل چندین نوع برچسب، به عنوان مثال ورزشی، سیاسی، اقتصادی، هنر و ... می‌باشد.

در این مقاله روشی پیشنهاد شده است که با ارائه ی الگوریتم های مورد نظر با استفاده از روش دمپستر شفر [۶] [۷] به نظرکاوی در تشخیص موضوع متون فارسی در شبکه های سایبر اجتماعی خواهیم پرداخت تا با استفاده از چالش های پیش رو به تحلیل و بررسی موضوع بپردازیم. در بخش دوم این مقاله، به بررسی روش های مشابه پرداخته و در بخش سوم روش پیشنهادی استفاده شده توضیح داده شده است. در بخش چهارم ، نتایج بیان گردیده و در نهایت در بخش آخر نتیجه گیری ارائه میشود.

۲. مروری بر ادبیات و کارهای مرتبط

عادل مجید و همکاران [۸] بر روی تشخیص احساس از متن به زبان اردو رومی متمرکز است. یک مجموعه جامع جمله ای توسعه داده شده که از حوزه های مختلف جمع شده و آن را با شش کلاس مختلف حاشیه نویسی می کنند. برای استخراج ویژگی از Word2Vec استفاده شده است . برای طبقه بندی از الگوریتم های مختلف پایه مانند کی-نزدیکترین همسایه ، درخت تصمیم ، ماشین بردار پشتیبان و جنگل تصادفی را بر روی مجموعه اعمال شد. پس از آزمایش و ارزیابی ، بر طبق ادعای این پژوهش مدل ماشین بردار پشتیبان نسبت به بقیه ی الگوریتم های طبقه بندی با دقت ٪۶۹,۵۴ به نتایج بهتری دست یافته است.

آزمین و همکاران [۹] شناسایی احساس چند کلاسه از متن بنگلا را پیشنهاد کردند که با استفاده از طبقه بندی کننده چند جمله ای بیز ساده همراه با ویژگی های مختلف مانند ریشه یابی ، برچسب گذاری قسمت های گفتار (POS)، انگرمس ، فرکانس اصطلاح معکوس فرکانس (TF-IDF) می باشد .با ادعای این پژوهش مدل نهایی توانست متن را در سه کلاس احساسی با دقت کلی ٪۷۸,۶ طبقه بندی کند.

رشیدی و همکاران [۶] روش های ادغام تصمیم گیری برای داده های سنجش از دور را پیشنهاد نموده اند . در همجوشی سطح تصمیم گیری ، نتایج دریافتی از طبقه بندی کننده محلی مختلف ترکیب می شود و تصمیم نهایی تعیین می شود. مهمترین روشهای همجوشی تصمیم گیری که در کاربردهای مختلف به کار رفته است و در این پژوهش نیز مورد استفاده قرار گرفته : تابع رای گیری اکثریت[۲] (MVF) ، بهترین روش رای گیری اکثریت[۱] (BMVF) ، روش مبتنی بر رتبه[۲]

[۱] Text classification
[۲] Majority Voting Function

و روش مبتنی بر امتیاز[۳] ، استنباط بیزی[۴] و روش دمپستر-شفر[۵] است. نتایج فیوژن با توجه به قابل قبول بودن و صحت طبقه بندی آنها با هم مقایسه می شود. باتوجه به ادعای این پژوهش نتایج نشان می دهد که روش دامپستر-شفر دقیق تر از نسبت به سایر روش های پیشنهادی با دقت ۸۸٪ است.

۳. روش پیشنهادی تشخیص موضوع متون خبری فارسی در شبکه های سایبر اجتماعی

مسئله ی مورد استفاده در این مقاله ، طراحی یک مدل کارا برای تشخیص متون خبری فارسی است. همانطور که در جمع بندی پژوهش های پیشین بیان کردیم روش ادغام تصمیم از دقت و عملکرد بهتری نسبت به روش های مستقل عمل می کنند. بنابراین می توان گفت پژوهش حاضر بر طبق همین روش خواهد بودچارچوب کلی روش پیشنهادی بر طبق روش دمپستر شفر خواهد بود که با استفاده از آن به بهبود نتایج پرداخته خواهد شد.

شکل ۱- چهارچوب روش پیشنهادی

روش پیشنهادی ارائه شده شامل بخش های زیر می باشد که در ادامه هر کدام از این موارد توضیح داده خواهد شد.

۱.۳. مجموعه داده

مجموعه داده‌ی همشهری بر اساس استاندارد TREC در گروه تحقیقاتی دانشگاه تهران تولید شده است [۱۰]. اسناد موجود در این مجموعه داده شده به برخی دسته‌های مختلف طبقه‌بندی شده و در قالب XML و استاندارد UTF-8 ذخیره شده می‌باشند. در این مقاله از ۵۰۰۰ سند از این مجموعه که شامل ۵ کلاس ورزشی، سیاسی، علمی، اقتصادی و اجتماعی می‌باشد، استفاده شده است. مجموعه‌ی اسناد به صورت تفکیک شده در جدول ۱ نشان داده شده‌اند.

---

[۱] Best Majority Voting Function
[۲] Rank Based Method
[۳] Score Based Method
[۴] Bayesian Inference Method
[۵] Dempster-Shafer Method (D-S) and Extended D-S

جدول ۱- تعداد اسناد هر موضوع در مجموعه‌ی داده

موضوع	تعداد اسناد در مجموعه داده
ورزشی	۱۰۰۰
اجتماعی	۱۰۰۰
اقتصادی	۱۰۰۰
علمی	۱۰۰۰
سیاسی	۱۰۰۰

از مجموعه‌ی ۵۰۰۰ سند، ۸۰ درصد آن برای داده‌ی آموزشی و مابقی برای داده‌ی تست استفاده شده که از هر موضوع ۲۰درصد انتخاب شده و در هر مرحله مورد استفاده قرار گرفته است.

۲.۳. پیش پردازش

یکی از مراحل مهم داده‌کاوی، پیش‌پردازش دادگان است. پیش‌پردازش، داده را به قالب مناسب برای داده‌کاوی تبدیل کرده و روند محاسبات و استخراج اطلاعات را تسریع و ساده می‌کند [۱۱].پردازش زبان فارسی از جهاتی با پردازش زبان انگلیسی تفاوت دارد. در زبان انگلیسی تمامی حروف و تمامی کلمات جدا از هم و با قانونی مشخص نوشته می‌شوند و این در حالی است که در زبان فارسی بعضی از حروف به هم چسبیده هستند، برخی از حروف جدا از هم نوشته می‌شوند، بعضی از کلمات یکپارچه‌اند، بعضی از کلمات با فاصله یا نیم‌فاصله به دو یا چند بخش تقسیم می‌شوند. تمامی حوزه‌های مرتبط با پردازش زبان طبیعی به نحوی با متون واقعی سروکار دارند.اگر حروف، نشانه‌های نگارشی و کلمات فارسی به شکل یکسانی نوشته نشوند، متون مورد استفاده قابل تحلیل توسط سامانه‌های رایانه‌ای نخواهند بود. به عنوان مثال اگر نرمال‌سازی روی دو داده‌ی "رئیس‌جمهور" و "رییس‌جمهور" اعمال نشود، سیستم این دو را دو عبارت جدا در نظر می‌گیرد که روی نتایج تأثیر زیادی دارد. طی فرایند نرمال‌سازی، علائم نگارشی، حروف، فاصله‌های بین کلمات، اختصارات و غیره بدون ایجاد تغییرات معنایی در متن به شکل استاندارد تبدیل می‌گردند. بنابراین، بایستی از یک استاندارد مشترک برای پیش‌پردازش و پردازش

متون استفاده کرد. همچنین کارکترها و کلمات زائد و توقف حذف شده و کاراکترها نرمال می‌شوند. به عنوان مثال، برای دو کلمه «مسئله» و «مسأله»، کل متن نرمال شده و یکی از این دو و یا یک کلمه جایگزین مانند «مساله» به عنوان کلمه مرجع انتخاب می‌شود [۱۲]. از این قبیل کلمات می‌توان «رییس» و «رئیس»، کلماتی که دارای حروف "ی" و "ک" عربی می‌شوند و ... را نام برد.همچنین فاصله بین کلمات و یا عبارات نرمال شده و همه به یک فاصله تبدیل می‌شوند. به عنوان مثال "میرفت"، "می‌رفت" و "می-رفت" هر سه دارای یک معنا و مفهوم هستند، اما اگر پردازش روی آن‌ها صورت نگیرد، دارای نتایج متفاوتی خواهند بود [۱۲].

۳.۳. استخراج ویژگی

پس از اتمام مرحله پیش پردازش ، استخراج ویژگی برای ارزیابی داده های پردازش شده اعمال می شود. انتخاب و استخراج ویژگی مهمترین مرحله برای تشخیص احساسات است زیرا بر نتیجه کلی کار تأثیر می گذارد. یک انتخاب ویژگی خوب منجر به پیش بینی خوبی می شود. بنابراین ، انتخاب صحیح ویژگی ها برای ارتقا طبقه بندی بسیار مهم است.

Unigram: (واژگان منفرد) در این مرحله ابتدا پاراگراف به جملات تقسیم می شوند و سپس جملات به صورت کلمه تبدیل می شوند.

Bigram : در این مرحله ابتدا پاراگراف به جملات تقسیم می شوند و سپس جملات به صورت واژگان دوتایی تبدیل می شوند.

TFIDF: هدف نشان دادن اهمیت کلمه کلیدی مورد نظر از طریق مقایسه تعداد تکرار کلمه در متن با تکرار آن کلمه در مجموعه ای بزرگ تر از مستندات می‌باشد.

Hashing Vectorizer: یک برداری است که از ترفند هش برای یافتن نام رشته نشانه برای مشخص کردن نگاشت شاخص اعداد صحیح استفاده می کند.تبدیل اسناد متنی به ماتریس توسط این بردار انجام می شود که در آن مجموعه اسناد را به یک ماتریس پراکنده تبدیل می کند که تعداد وقوع نشانه ها را در خود نگه می دارد.

Text To Sequences: این روش یک روش مبتنی بر جمله است که برای هر کلمه یک عدد در نظر گرفته می شود و سپس طول جمله های کوتاه با صفر پر می شود تا طول جمله ها با یکدیگر مساوی شوند تا بدین ترتیب یک ماتریس به اندازه n سند و mویژگی بدست آید.

۴،۳. انواع طبقه بندی یادگیری ماشین

الگوریتم مورد استفاده از نوع الگوریتم ماشین بردارپشتیبان (SVM) ،بیز ساده (NB) و کی نزدیکترین همسایه (KNN) و درخت تصمیم (DT) و جنگل تصادفی (RF) که از الگوریتم های نظارت شده در یادگیری ماشین هستند استفاده شده که در نهایت با استفاده از روش دمپستر شفر نتایج بهبودی بدست می آید .

۵.۳. روش دمپسترشفر

در ادغام تصمیم گیری ، نتایج دریافتی از طبقه بندی کننده مختلف ترکیب می شود و تصمیم نهایی تعیین می شود.دمپستر برای اولین بار نظریه شواهد دمپستر-شفر را که به عنوان نظریه توابع باور نیز شناخته می شود، در سال ۱۹۶۷ معرفی کرد و شاگردش شفر آن را برای اولین بار در سال ۱۹۷۶ گسترش داد. [۱۳] [۱۴،۱۵]

این به عنوان تعمیم نظریه بیزی در نظر گرفته می شود. مشابه با نظریه استنتاج بیزی، روش D-S یک تابع جرم پیشینی را برای به دست آوردن یک بازه شواهد پسینی به روز می کند.فاصله اثباتی اعتبار (اندازه گیری اعتقاد) یک گزاره و معقول بودن آن را کمیت می کند.

فرض کنید $\theta = \{A_1, \dots, A_N\}$ که به عنوان چارچوب تشخیص شناخته می شود ، مجموعه محدودی از گزاره های N درمورد یک موضوع موضوعی منحصر به فرد و کامل است. بنابراین مجموعه توان θ (که به عنوان 2^θ مشخص می شود) ، متشکل از همه زیر مجموعه های θ ، به شرح زیر است:

$$2^{\theta} = \{ A_1, \dots, A_N, A_1 \cup A_2, \dots, A_1 \cup A_N, \dots \} \quad (۱)$$

که در آن U عملگر اتحادیه مجموعه ها است.

روش DS ، به جای اختصاص احتمال به فرضیه ها (روش بیزی) ، توده های احتمال $m(A_i)$ را به دو گزاره منفرد و ترکیبی اختصاص می دهد. تابع جرم احتمال به صورت زیر تعریف می شود:

$$m : 2^{\theta} \to [0,1] \quad , \quad \sum_{B_i \subset 2^{\theta}} m(B_i) = 1 , \quad m(\emptyset) = 0 \quad (۲)$$

جایی که φ مجموعه خالی است و ⊂ عملگر زیر مجموعه است. احتمال یک گزاره A_i با جمع کردن جرم‌های احتمال برای θ و 2^{θ} عناصر مرتبط به دست می‌آید

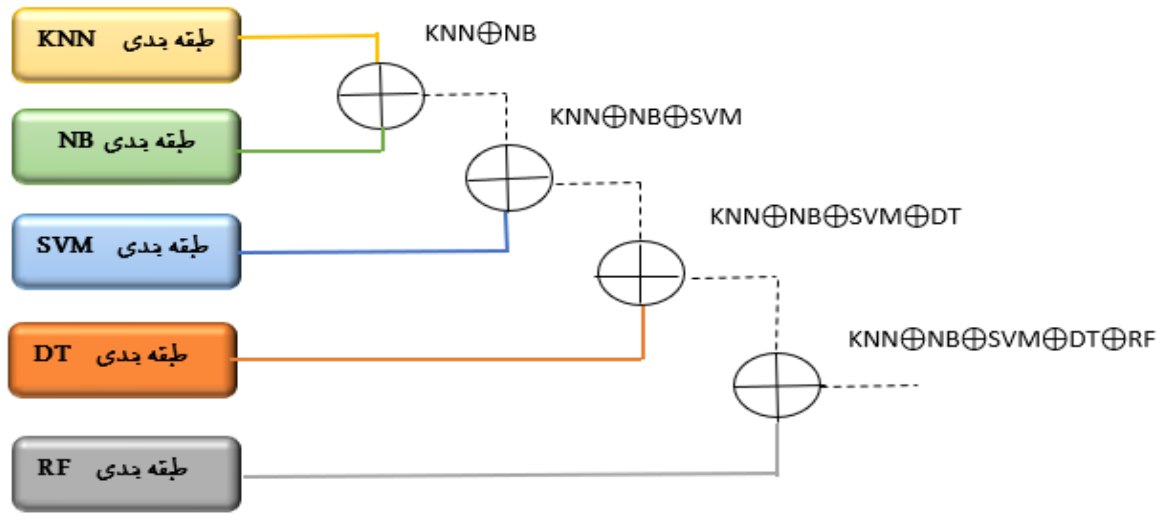
$$prob.(A_i) = \sum_{B_i \in \theta, 2^{\theta}} m(B_j) , \quad A_i C B_j \quad (۳)$$

قانون DS برای دو منبع مستقل را می توان به صورت زیر نوشت:

$$f : f(m_1(A_i), m_2(B_j)) \in R^+ \quad (۴)$$

$$m(u_l) = \frac{\sum_{A_i \cap B_j = u_l} f(m_1(A_i)\, m_2(B_j))}{\sum_{A_i \cap B_j \neq \emptyset} f(m_1(A_i)\, m_2(B_j))}$$

که در آن ∩ تقاطع مجموعه ها را نشان می دهد و m_i تابع احتمال (طبقه بندی کننده) است. علاوه بر این u_lگزاره ای است که به عنوان ترکیبی از فرضیه های اساسی ، A_i و B_j تعریف می شود.



شکل ۲ – شمای کلی قاعده ی ترکیب روش دمپستر شفر در چهارچوب روش پیشنهادی

روش دمپسترشفر دارای ویژگی خاصیت جابه جایی پذیری و انجمنی است. به این معنا که نه ترتیب ورود گزاره های مختلف و نه گروه بندی های متفاوت شواهد تاثیری در نتیجه ی ترکیب طبق قاعده ی ترکیب یا حاصل جمع متعادل نخواهد داشت.

۴. معیارهای ارزیابی

در مواردی که دادگان برچسب دار یا حاشیه نویسی شده از نظرات موجود است، چهار معیار درستی ، دقت ،بازخوانی ، ضریبf برای ارزیابی کارایی الگوریتم ها مطرح است.
درستی یعنی درصد کل مشاهداتی است که به درستی طبقه بندی شده اند. درستی به صورت زیر محاسبه میشود:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (۵)$$

TP شامل شناسایی های درست، TN شامل حذف های درست، FP شناسایی نادرست و FN حذف های نادرست است.
به نسبتِ بین تمام مشاهداتی که به درستی طبقه بندی شده اند (TP)به تمام مشاهدات طبقه بندی شده مثبت (FP+TP) ،دقت گفته می شود.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (۶)$$

بازخوانی ، نسبتِ بین مشاهدات به درستی طبقه بندی شده به تمام مشاهدات مثبت است. به این معیار، حساسیت و یا نرخ تأیید درست هم گفته می شود.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (۷)$$

ضریب f در واقع میانگین هارمونیک بین بازخوانی و دقت است. برای ارزیابی نتایج مدل میتوان از این ضریب (score-f) استفاده کرد و برای محاسبه ی آن از دو پارامتر دقت و بازخوانی استفاده می شود.

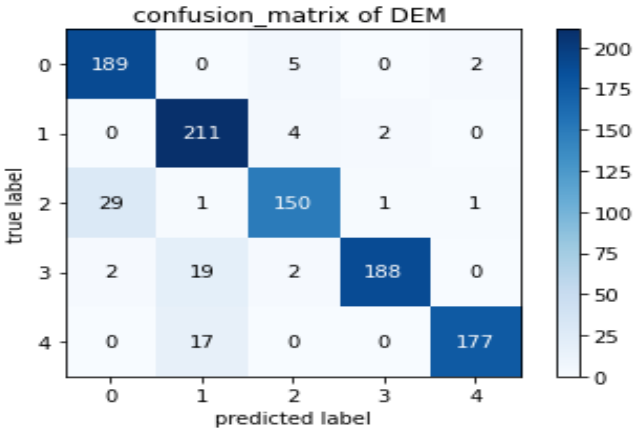
$$\text{f} - \text{score} = 2 * \frac{\text{Precision} \,.\, \text{Recall}}{\text{Precision+ Recall}} \quad (۸)$$

بررسی ویژگی های در متن فارسی و ارزیابی تأثیر هر گروه از ویژگی ها نیازمند تحقیقات متعددی است و در این تحقیق با توجه به اهداف آن مجموعه ای از این ویژگی های متنی مورد استفاده در (جدول ۲) در نظر گرفته شده است.

جدول ۲ - لیست ویژگی های مورد استفاده در روش دمپسترشفر

ادغام تصمیم گیری	ویژگی	ضریب f	بازخوانی	دقت	درستی
روش دمپسترشفر	Unigram	۰,۸۷	۰,۸۷	۰,۸۸	۰,۹۱
	Bigram	۰,۸۵	۰,۸۳	۰,۸۸	۰,۹۰
	TFIDF	**۰,۸۷**	**.۰,۸۹**	**۰,۸۵**	**۰,۹۲**
	Hashing Vectorizer	۰,۸۲	۰,۷۸	۰,۹۱	۰,۸۹
	Text To Sequences	۰,۸۴	۰,۸۳	۰,۸۷	۰,۸۹

در حوزه ی هوش مصنوعی، ماتریس سردرگمی[۱] به ماتریسی گفته می‌شود که در آن عملکرد الگوریتم‌های مربوطه را نشان می‌دهند. برای گسترش روش دمپستر شفر نیز ، ماتریس سردرگمی به دست آمده از نتایج طبقه بندی در نظر گرفته می شود.



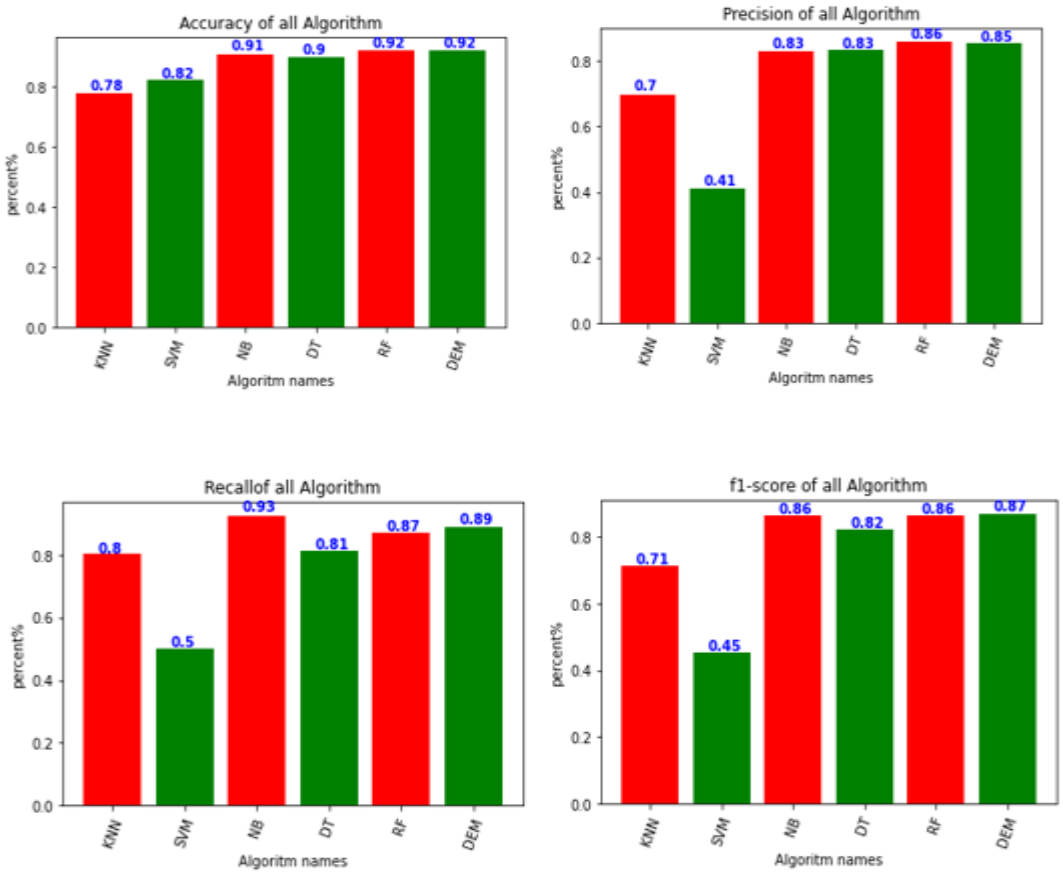
شکل ۳- ماتریس سردرگمی دمپسترشفر

۵. نتیجه گیری

در این مقاله ،یکی از روش های ادغام تصمیم گیری به نام روش دمپستر شفراتخاذ شده است که مقایسه ای بین طبقه بندی کننده های SVM ، KNN ، NB و DT و RF در مورد نظرکاوی در سطح جمله برای تشخیص موضوع متون خبری در پنج کلاس (سیاسی ، فرهنگی ، ورزشی ، علمی ، اقتصادی) انجام شده است. پژوهش صورت گرفته به تحلیل موضوع بر روی زبان فارسی بوده که از نقطه قوت آن بشمار می رود. در این تحقیق جهت تقویت این موضوع از ادغام تصمیم با هدف افزایش دقت تشخیص موضوع استفاده شده است. در نهایت نتایج حاصل از انواع طیقه بندی با استفاده از روش‌های ادغام

[۱] Confusion Matrix

تصمیم مانند روش دمپسترشفر ترکیب شدند. روش پیشنهادی با استفاده از مجموعه داده فارسی همشهری پیاده سازی و ارزیابی شده است. نتایج آزمایشات این پژوهش نشان می‌دهد که در مقایسه با روش‌های قبلی اجرا شده برروی همین مجموعه داده، استفاده از روش پیشنهادی باعث افزایش معیار صحت با دقت ۹۲٪ در تشخیص موضوع می‌شود. با توجه به پژوهش انجام شده، پیشنهادهایی برای پژوهش های آتی در این زمینه را می توان با سایر روش های ادغام تصمیم ارائه نمود.و همچنین چون دقت طبقه بندی ها به انواع استخراج ویژگی وابستگی شدیدی دارد. انتخاب و استخراج ویژگی مهمترین مرحله است زیرا بر نتیجه کلی کار تأثیر می گذارد توصیه می شود از استخراج ویژگی های متنوع همچون Word2vec و TF-IDF وزن دار نیز استفاده شود.



شکل ۴- مقایسه صحت ارزیابی روش دمپسترشفر با انواع طبقه بندی یادگیری ماشین

۶. مراجع

[1] Sebastiani, F., Machine learning in automated text categorization. ACM computing surveys (CSUR), 2002. 34(1): p. 1-47.

[2] M. Fernández-Gavilanes, J. Juncal-Martínez, S. García-Méndez, E. Costa-Montenegro, and F. J. González-Castaño, “Creating emoji lexica from unsupervised sentiment analysis of their descriptions,” Expert Syst. Appl., vol. 103, no. August, pp. 74–91, 2018, doi: https://doi.org/10.1016/j.jocs.2017.01.010.

[3] Y. Zhang, D. Song, P. Zhang, P. Wang, X. Li, J., Li, and B. Wang, “A quantum-

inspired multimodal sentiment analysis framework," Theor. Comput. Sci., vol. 752, no. December, pp. 21–40, 2018, doi: https://doi.org/10.1016/j.tcs.2018.04.029.

[4] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," Inf. Sci. (Ny)., vol. 311, no. August, pp. 18–38, 2015, doi: https://doi.org/10.1016/j.ins.2015.03.040.

[5] A. A. Lazarus and C. N. Lazarus, "The 60-second shrink," Atascadero, CA. 1997.

[6] Ali Rashidi , Hassan Ghassemian , EXTENDED DEMPSTER-SHAFER THEORY FOR MULTI-SYSTEM/SENSOR DECISION FUSION , Commission IV, Working Group IV/7.

[7] Morteza Zangeneh Soroush1 , Keivan Maghooli1, Seyed Kamaledin Setarehdan and Ali Motie Nasrabadi ,A novel approach to emotion recognition using local subset feature selection and modifed Dempster-Shafer theory (2018).

[8] Adil Majeed , Hasan Mujtaba , Mirza Omer Beg , Emotion Detection in Roman Urdu Text using Machine Learning, 2020 35th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW).

[9] Sara Azmin , Kingshuk Dhar , Emotion Detection from Bangla Text Corpus Using Naïve Bayes Classifier , 4th International Conference on Electrical Information and Communication Technology (EICT), 20-22 December 2019, Khulna, Bangladesh.

[10] AleAhmad, A., et al., Hamshahri: A standard Persian text collection. Knowledge-Based Systems, 2009. 22(5): p. 382-387.

[11] Saraswathi, M. and V. Balu, Preprocessing techniques for effective data extraction and computation. IUP Journal of Computer Sciences, 2013. 7(3): p. 27.

[12] Shamsfard, M., H.S. Jafari, and M. Ilbeygi. STeP-1: A Set of Fundamental Tools for Persian Text Processing. in LREC. 2010

[13] Hall D.L., 1992. Mathematical Techniques in Multi-Sensor Data Fusion, Artech House, Inc.

[14] Hongwei Z., Basir O. and Karray F., 2002. Data fusion for pattern classification via the dempster-shafer evidence theory. IEEE International Conference on Systems, Man and Cybernetics, 7, pp.109 -110.

[15] Foucher S., Germain M., Boucher J. M. and Benie G.B., 2002. Multisource classification using ICM and Dempster-Shafer theory. IEEE Transactions on Instrumentation and Measurement, 5, pp. 277 -281.