دانشگاه آزاد اسلامی واحد بندرعباس

فرم طرح تحقیق (پروپوزال)

درخواست تصویب موضوع پایان نامهٔ

کارشناسی ارشد

نام و نام خانوادگی دانشجو:

رشته/گرایش: مهندسی کامپیوتر – نرم افزار

عنوان تحقیق (پروپوزال):

تشخیص نویسنده ی یک متن با استفاده از یادگیری ماشین و ترکیب طبقه بندها

تاریخ تحویل فرم به دفتر گروه تخصصی:
امضاء دانشجو:

توجه:

1- فرم بایستی تایپ شده تحویل گردد.

2- سایر نکات در تکمیل فرم از طریق سایت و دفتر گروه آموزشی دریافت گردد.

امضااستادراهنما

فرم طرح تحقیق(پروپوزال)

این قسمت توسط حوزهٔ معاونت پژوهشی دانشگاه پر می‌شود .

شماره :
تاریخ :
پیوست :

دانشگاه آزاد اسلامی واحد بندرعباس

فرم طرح تحقیق(پروپوزال)

درخواست تصویب موضوع پایان نامهٔ کارشناسی ارشد

توجه: این فرم با هدایت استاد راهنما و مشاور تکمیل شود .

عنوان تحقیق به فارسی:

تشخیص نویسنده ی یک متن با استفاده از یادگیری ماشین و ترکیب طبقه بندها

عنوان تحقیق به انگلیسی:

Detecting the author of a text using machine learning and classifiers composition

کلمات کلیدی پایان نامه:

1- شناسایی نویسنده2- الگوریتم شبکه بیزین3-الگوریتم بردار پشتیبان خطی 4- الگوریتم رگرسیون منطقی 5- الگوریتم درخت های اضافی 6- ترکیب دسته بندها 7-رای گیری اکثریت 8- انتخاب روبه جلو 9- حذف رو به عقب 10- یادگیری تجمعی

1- **اطلاعات مربوط به دانشجو**

نام: **نام خانوادگی:**

شماره دانشجویی:
رشته/گرایش:مهندسی کامپیوتر – نرم افزار
سال ورود: **مهر1397 بهمن.....**
نشانی پستی:
تلفن:

2- اطلاعات مربوط به استاد راهنما

نام: عباس نام خا نوادگی:عکاسی نام پدر:

شماره شناسنامه: کدملی : سال تولد : محل تولد :

محل صدور: استان شهرستان:تبریز

آخرین مدرک تحصیلی دانشگاهی / حوزوی:

سنوات تدریس کارشناسی ارشد/دکترا: تخصص اصلی: تخصص جنبی:

پایه: مرتبه علمی(رتبه دانشگاهی):

نحوه همکاری: تمام وقت□ نیمه وقت□ مدعو□

نشانی پستی:

تلفن:

الف) تعداد پایان نامه ها / رساله‌های راهنمایی شدهٔ در یک سال گذشته:

ردیف	عنوان پایان نامه	مقطع	نام دانشگاه	سال
1				
2				
3				
4				
5				
6				
7				

ب) تعداد پایان نامه ها / رساله‌های در دست راهنمایی:

ردیف	عنوان پایان نامه	نام دانشگاه	سال
1			
2			
3			
4			
5			
6			
7			

تذکر: براساس بخشنامه سازمان مرکزی دانشگاه آزاد اسلامی ،ظرفیت اساتید تمام وقت واحد بندرعباس 5 دانشجو ،اساتید نیمه وقت و مدعو 3 دانشجو می باشد.لذا خواهشمند است موضوع ظرفیت ها رعایت گردد.

امضا استاد راهنما

3- اطلاعات مربوط به استاد مشاور (دانشجوزمانی زمانی می‌تواند از مشاور استفاده نماید که تائیدیه کتبی آن را قبلاً از گروه تخصصی اخذ و به پیوست پروپوزال ارئه نماید.)

نام: نام خا نوادگی: نام پدر:

شماره شناسنامه: کدملی : محل صدور: استان شهرستان

آخرین مدرک تحصیلی دانشگاهی / حوزوی:

سنوات تدریس کارشناسی ارشد/دکترا: تخصص اصلی: تخصص جنبی:

پایه: مرتبه علمی(رتبه دانشگاهی):

نحوه همکاری: تمام وقت□ نیمه وقت□ مدعو□

نشانی پستی:

تعداد پایان نامه‌ها / رساله های در دست مشاوره:

ردیف	عنوان پایان نامه	مقطع	نام دانشگاه	سال
1				
2				
3				
4				
5				

تذکر: براساس بخشنامه سازمان مرکزی دانشگاه آزاد اسلامی ،ظرفیت اساتید تمام وقت واحد بندرعباس 5 دانشجو ،اساتید نیمه وقت و مدعو 3 دانشجو می باشد.لذا خواهشمند است موضوع ظرفیت ها رعایت گردد.

امضا استاد مشاور

4- اطلاعات مربوط به پایان نامه:

الف) فارسی ☐

عنوان فارسی :

تشخیص نویسنده ی یک متن با استفاده از یادگیری ماشین و ترکیب طبقه بندها

ب) انگلیسی ☐

عنوان انگلیسی :

Detecting the author of a text using machine learning and classifiers composition

پ: نوع کار تحقیقاتی: بنیادی ☐ نظری ☐ کاربردی ■ عملی ☐

ت: تعداد واحد پایان نامه:

ث: کلمات کلیدی پایان نامه: 1- شناسایی نویسنده2- الگوریتم شبکه بیزین 3-الگوریتم بردار پشتیبان خطی 4- الگوریتم رگرسیون منطقی 5- الگوریتم درخت های اضافی 6- ترکیب دسته بندها 7-رای گیری اکثریت 8- انتخاب روبه جلو 9- حذف رو به عقب 10- یادگیری تجمعی

ج: سئوال اصلی تحقیق :تاثیر یادگیری تجمعی در بهبود نتایج نسبت به الگوریتم های دسته بندی چگونه می باشد ؟

a. تحقیق بنیادی پژوهشی است که به کشف ماهیت اشیاء پدیده‌ها و روابط بین متغیرها ، اصول ، قوانین و ساخت یا آزمایش تئوری‌ها و نظریه‌ها می‌پردازد و به توسعه مرزهای دانش رشتهٔ علمی کمک می‌نماید .

b. تحقیق نظری: نوعی پوشش بنیادی است و از ارزش استدلال و تحلیل عقلانی استفاده می‌کند و بر پایهٔ مطالعات کتابخانه‌ای انجام می‌شود .

c. تحقیق کاربردی: پژوهشی است که با استفاده از نتایج تحقیقات بنیادی به منظور بهبود و به کمال رساندن رفتارها ، روشها ، ابزارها ، وسایل ، تولیدات ، ساختارها و الگوهای مورد استفاده جوامع انسانی انجام می‌شود .

d. تحقیق علمی: پژوهشی است که با استفاده از نتایج تحقیقات بنیادی و با هدف رفع مسائل و مشکلات جوامع انسانی انجام می‌شود .

5- بیان مسأله (حداقل یک صفحه با ذکر منبع معتبر علمی شامل تشریح ابعاد ، حدود مسأله ، معرفی دقیق مسأله ، بیان جنبه‌های مجهول و مبهم و متغیرهای مربوط به پرسش‌های تحقیق ، منظور تحقیق،)

رویکردهای کمّی به وظایفی مانند تخصیص دادن اسناد ، تایید، ثبت مشخصات و یا خوشه‌بندی نویسنده بر این فرض پایه تکیه دارند که سبک نوشتاری اسناد به گونه‌ای مشخص ، آموخته و برای ساختن مدل‌های پیش‌بینی استفاده می‌شود . هر سیستم شناسایی موفقِ نویسنده ، می‌تواند در یک تنظیمات اختصاصی یا در یک تنظیمات تایید(بازبینی) ، تعیین هویتِ قدرتمندی در متن در ژانرهای مختلف ، رفتار با موضوعات مختلف و یا داشتن مخاطبان هدف مختلف در ذهن ایجاد کند . مدل‌های سنتی برای تخصیص دادن اسناد در مواردی که چندین نویسنده در یک سند مشارکت دارند، قابل‌اجرا نیستند . در نتیجه می توان با ابزار های داده کاوی این موضوع را حل کرد .

در این پژوهش ابتدا با توجه به ماهیت مسئله با استفاده از کتابخانه پاندا داده های ورودی با سه ویژگی را دریافت می کنیم . روش دسته بندی یک روش یادگیری با نظارت است که داده‌های ورودی به سه بخش داده‌های آموزش و داده‌های آزمون و داده های اعتبارسنجی تقسیم می‌شوند . هر الگوریتم کاندید، ابتدا با استفاده از مجموعه داده آموزش یک مدل را که نشان دهنده الگوی حاکم بر داده‌ها می‌باشد را استخراج می‌کند و سپس با استفاده از مجموعه آزمون و اعتبارسنجی، دقت مدل ارائه شده برای دسته‌بندی را بررسی می‌کند.

امضا استاد راهنما

الگوریتم‌های متعددی برای دسته بندی ارائه شده‌اند که از آن دسته می‌توان؛ به شبکه‌های بیزین، الگوریتم دسته بندی کننده بردار پشتیبان خطی ، الگوریتم رگرسیون منطقی والگوریتم درخت های اضافی اشاره کرد.
در این نوشتار ابتدا با استفاده از توابع کتابخانه پاندا داده های مربوط به نویسندگان تقسیم بندی می¬شوند تا به صورت داده های آموزش و آزمون و اعتبار سنجی درآیند . سپس بعد از پاکسازی داده و تهیه ورودی برای الگوریتم های پایه که در بالا ذکر شد، 100 دسته بند را با استفاده از هر الگوریتم میسازیم . سپس با استفاده از معیار های ارزیابی بر روی داده های اعتبار سنجی، هر مدل را ارزیابی کرده و بهترین دسته بند را انتخاب میکنیم و در مرحله بعد با روش های ترکیبِ دسته بند ها از جمله حذف پس رونده و انتخاب پیش رونده و انتخاب تک دسته بندِ بهترین، بهترین روش را انتخاب کرده و بر روی داده های آزمون تست می¬کنیم . در نهایت سعی شده با در نظر گرفتن نقاط ضعف و قوت روش‌های مختلف داده کاوی یک الگوریتم ترکیبی برای تشخیص نویسنده ارائه شود.

6- سوابق مربوط (بیان مختصر سابقهٔ تحقیقات انجام شده دربارهٔ موضوع و نتایج به دست آمده در داخل و خارج از با ذکر حد اقل 10 سابقه مطالعات علمی مرتبط با ذکر منابع)

امضا استاد راهنما

7- فرضیه‌ها: (هر فرضیه به صورت یک جملهٔ خبری نوشته شود- حداقل 2 فرضیه به مرحله آزمون گذاشته می‌شود)

- یادگیری تجمعی بهبود موثری در نتایج تشخیص نویسنده متن نسبت به الگوریتم های دسته بندی دارد.
- استفاده از رویکرد فراوانی کلمه کلیدی (TF-IDF) نتایج بهتری در مقایسه با استفاده از کیسه کلمات با فرکانس کلمات (BOW) و کیسه کلمات دودویی (BOW_binary) دارد.
- استفاده از میانگین معیارهای ارزیابی مدل های ساخته شده به عنوان حد آستانه جهت انتخاب مدل ها می تواند نتایج بهتری را حاصل نماید.

8- اهداف تحقیق (شامل اهداف علمی ، کاربردی و ضرورت‌های خاص انجام تحقیق- حداقل 2 هدف اصلی در ارتباط با فرضیه‌ها و موضوع تحقیق)

1-هیچ راهنمای مشخصی برای آنکه پژوهشگران یا تحلیلگران بدانند چگونه یک الگوریتم داده کاوی را انتخاب کنند وجود ندارد. در نتیجه انتخاب یک الگوریتم مشخص امری بسیار پیچیده است، لذا در این پژوهش برای ارتقای نتایج داده کاوی از چندین الگوریتم استفاده شده و هر کدام را جداگانه توضیح و تشریح داده و پردازش‌ها را با الگوریتم‌های مختلف تکرار و سرانجام از تکنیک رای گیری اکثریت استفاده می شود. هدف از روش های ارائه شده افزایش دقت تشخیص نویسنده با استفاده ازکشف الگوها ، درمیان مجموعه داده ها می‌باشد.

2- بررسی میزان تاثیر استفاده از رویکرد فراوانی کلمه کلیدی (TF-IDF) در مقایسه با استفاده از کیسه کلمات با فرکانس کلمات (BOW) و کیسه کلمات دودویی (BOW_binary)

3- بررسی میزان تاثیر به کارگیری میانگین معیارهای ارزیابی مدل های ساخته شده تحت عنوان حد آستانه ، جهت اتنخاب مدل ها

4- ساختار مجموعه داده موجود، نتایج مورد انتظار در خروجی، شناخت داده‌کاو از یک الگوریتم و مولفه‌های پیکربندی پایگاه داده در انتخاب الگوریتم مناسب داده‌کاوی تاثیرگذار

9- در صورت داشتن هدف کاربردی بیان نام بهره‌وران (اعم از موسسات آموزشی و اجرایی و غیره):

1	سازمان های اطلاعاتی	6	
2	موسسات علمی و مراکز تحقیقاتی مرتبط	7	
3	دانشجویان رشته مهندسی نرم افزار و هوش مصنوعی	8	

امضا استاد راهنما

10- جنبهٔ نوآوری و جدید بودن تحقیق در چیست ؟ (این قسمت توسط استاد راهنما تکمیل شود)

- استفاده از روش های ترکیب دسته بندها
- ساخت تعداد مدل های متعدد
- استفاده از میانگین معیارهای ارزیابی مدل های ساخته شده یه عنوان حد آستانه جهت انتخاب مدل ها می تواند نتایج بهتری را حاصل نماید.

امضاء استاد راهنما

11-روش کار:

1- خواندن مجموعه داده

اولین قدم برای آغاز کار هوش مصنوعی، بارگذاری داده است.

2- تجزیه و تحلیل داده

از مهم ترین قسمت های کار با داده ، شناخت داده ورودی است . تسلط بر اینکه داده ورودی، شامل چه آیتم هایی است ، چه ویژگی و چه شیء هایی در داده وجود دارد، شامل چند شیء است و نوع ویژگی و اشیاء چیست . همچنین بررسی برچسب ها ، نوع آن ها و تعداد برچسب تاثیر بسیاری در نحوه کار با داده و انتخاب روش های مختلف کار با داده دارد .

داده استفاده شده در این پژوهش ، شامل 19579 شیء و سه ویژگی می باشد .

به عنوان ویژگی ،داده هایی نظیر نام نویسنده، ID و بخشی از متن کتاب قرار گرفته است . کل داده شامل 19579 جمله از کتاب های این سه نویسنده است . از سه نویسنده که با برچسب های EAP , MWS , HPL شناسایی میشوند ، تعداد 7900 متن از EAP و 6044 متن از MWS و 5635 متن از HPL در داده قرار داده شده است . و برای هر متن یک ID خاص در نظر گرفته شده است .

3- تقسیم بندی دیتاست

در این پژوهش تقسیم بندی دیتاست بر اساس سه دسته انجام گرفته است .

دسته اول داده آموزش28 است که که 70 درصد از داده اصلی را شامل میشود.دسته دوم داده آزمون29 است که 20 درصد از داده اصلی را شامل می شودو دسته سوم به عنوان داده اعتبارسنجی30 شناخته می شود که با نام اختصاری Dev تعریف شده و شامل 10 درصد از داده اصلی می باشد .

در واقع در این روش پس از یادگیری از روی داده آموزش و انجام آزمون روی داده آزمون، ارزیابی بر روی داده اعتبارسنجی انجام میگیرد.

مهم ترین فاکتور در ساخت این سه دسته، رعایت توزیع کلاس هاست . به نحوی که وقتی از داده اصلی، 70 درصد را به عنوان

امضا استاد راهنما

داده آموزشی استفاده می کنیم، توزیع کلاس ها یا برچسب ها باید به همان میزانی باشد که در داده اصلی موجود بوده است .

4- پاک¬سازی داده

پاک‌سازی داده‌ها یا تمیز کردن داده‌ها فرایند پیدا کردن، اصلاح کردن (یا حتی حذف کردن) داده‌های بی ارزش و اشتباه از دادگان یا پایگاه داده است. فرایند تمیز کردن داده‌ها ممکن است که از طریق ابزارهای داده‌کاوی یا پردازش دسته‌ای از طریق اسکریپت‌ها انجام شود. بعد از پاکسازی، مجموعه داده باید با سایر مجموعه داده‌های مشابه در سیستم سازگار باشد. ناسازگاری داده‌ها شناسایی و حذف (اصلاح) شده ممکن است بر اثر اشتباه انسانی هنگام ورود اطلاعات، انحراف در هنگام انتقال و ذخیره سازی اطلاعات یا به دلیل واژه نامه‌های داده مختلف باشد.

در این پژوهش برای پاکسازی داده به انجام عملیات نشانه گذاری ، حذف علایم ، کوچک کردن تمام حروف ، حذف کلمات کلیدی و یافتن ریشه لغات بسنده کرده ایم .

تهیه ورودی برای الگوریتم ها با استفاده از رویکردهای فراوانی کلمه کلیدی (TF-IDF)و کیسه کلمات با فرکانس کلمات (BOW)و کیسه کلمات دودویی(BOW_binary)

5- ساخت 100 نمونه داده آموزش با رویکرد جایگشتی

برای آموزش هوش مصنوعی ، بهتر است تعداد داده های بیشتری به عنوان ورودی به دسته بند های متفاوت داده شود .

با توجه به اینکه داده های دیتاست دارای محدودیت هستند، می توان با استفاده از همان داده ها، تعداد بیشتری داده آموزش ساخت . توجه داشته باشید که این روش بصورت جایگشتی بوده و با توجه به اینکه داده ها بصورت رندوم انتخاب می شوند ممکن است داده تکراری داشته باشیم . این روش را sampling with replacement میگویند .

6- انتخاب الگوریتم های پایه و ساخت 100 دسته بند با استفاده از هر الگوریتم

الگوریتم، روشی که برای جستجوی الگو در داده‌ها مورد استفاده قرار می‌گیرد را تعیین می‌کند و در واقع مانند یک روال ریاضی برای حل یک مساله خاص است. الگوریتم‌های گوناگونی برای تحلیل داده موجود هستند و لذا انتخاب الگوریتم داده‌کاوی مناسب یک مساله، برای پژوهشگران و تحلیلگران کاری دشوار است. هیچ راهنمای مشخصی برای آنکه پژوهشگران یا تحلیلگران چگونه الگوریتم انتخاب کنند وجود ندارد . انتخاب یک الگوریتم مشخص امری بسیار پیچیده است، لذا در این پژوهش برای ارتقای نتایج داده‌کاوی از چندین الگوریتم استفاده شده و پردازش‌ها را با الگوریتم‌های مختلف تکرار و دست آخر از از تکنیک رای گیری اکثریت استفاده می¬شود.

گاه نیاز به استفاده از چندین الگوریتم برای حل یک مساله واحد جهت حل فازهای مختلف مساله است. در مجموع می‌توان گفت هدف مساله، ساختار مجموعه داده موجود، نتایج مورد انتظار در خروجی، شناخت داده‌کاو از یک الگوریتم و مولفه‌های پیکربندی پایگاه داده در انتخاب الگوریتم مناسب داده‌کاوی تاثیرگذار هستند .

الگوریتم¬های پایه مورد استفاده در این مقاله شامل SVC خطی، بیز ساده، رگرسیون منطقی و درختان اضافی می باشد. همه این الگوریتم¬ها با 3 رویکرد BOW_Binary، BOW_TF و TF-IDF مدل¬سازی شده¬اند که به شرح و تفصیل هر یک می پردازیم.

7- انتخاب بهترین دسته بند

انتخاب بهترین دسته بند با مقایسه نتایج آن ها

امضا استاد راهنما

8- رای گیری اکثریت

هر مدل یک پیش‌بینی (رای‌گیری) برای هر نمونه آزمایش و پیش‌بینی خروجی نهایی همان چیزی است که بیشترین آرا را دریافت می‌کند .

9- انتخاب رو به جلو

انتخاب رو به جلو یک نوع رگرسیون گام به گام است که با یک مدل خالی شروع می شود و طبقه بندهایی را که به صورت نزولی بر مبنای یکی از معیارهای ارزیابی مرتب شده اند را یک به یک اضافه می کند و بهترین متغیر ، با برخی از معیارهای از پیش تعیین شده تعیین می شود و به مدل اضافه می گردد. معیار استفاده شده برای تعیین اینکه کدام مدل اضافه شود ، متفاوت است. در هر مرحله رو به جلو، شما یک طبقه بند را اضافه می کنید.در اینجا برای انجام مقایسات،Fscore ها مدنظر قرار داده شده اند.

10- حذف رو به عقب

انتخاب ویژگی به عقب مرتبط است ، و ممکن است با انتخاب کل مجموعه مدل¬ها شروع می‌شود و از آنجا به عقب کار می‌کند ، ویژگی‌هایی برای پیدا کردن زیر مجموعه بهینه از یک اندازه از پیش تعریف‌شده را حذف می‌کند .

11- ترکیب همه رویکردها:

در این رویکرد پیش بینی تمام رویکردها با هم جمع و بر روی این پیش¬بینی¬ها و برچسب مجموعه داده رای¬گیری اکثریت انجام گرفته است.

الف . شرح کامل روش کار و انجام این تحقیق (در این قسمت کلیه مراحل انجام تحقیق با جزئیات و ذکر منابع، روش های نمونه برداری، مطالعه و .... نوشته شود)

ب: روش گردآوری اطلاعات (میدانی ، کتابخانه ای و غیره):

روش گردآوری اطلاعات بصورت کتابخانه ای و تجربه (آزمایش و بررسی) خواهد بود. پس از انجام مطالعات کتابخانه ای و اینترنتی، مباحث مشابه و تکنیک های موجود مرتبط بر روی موضوع،تحقیق جمع آوری شده است.

پ: ابزار گردآوری اطلاعات:

مقالات ، نشریه ها و ژورنال های بین المللی، وب سایت های مجازی و کارهای انجام شده مربوط به تحقیق.

امضا استاد راهنما

ت: روش تجزیه و تحلیل اطلاعات و روشهای آماری(کامل توضیح داده شود):

در این مرحله مدل ها ی ایجاد شده مقایسه وارزیابی میشوند. معیارهای ارزیابی شامل دقت پیش بینی براساس معیار F1_score می باشد. در واقع F1-measure یک نوع میانگین بین پارامترp (دقت) و پارامتر(r یادآوری) است p .دقت سیستم در میان دادهای پیش بینی شده است r .نسبت تعداد داده های پیش بینی شده، به تعداد کل داده های مورد انتظار برای پیش بینی است. همچنین اطلاعات بدست آمده در جدول ذخیره و جهت مقایسه بین مدل های ساخته شده مورد استفاده قرار خواهد گرفت.

- درست مثبت : (TP)
- نادرست مثبت : (FP)
- نادرست منفی : (FN)
- درست منفی : (TN)

$$\text{بازخوانی (Recall)} = \frac{\text{تعداد های نمونه تشخیصی درست مثبت}}{\text{کل های نمونه واقعاً مثبت}} = \frac{TP}{TP+FN}$$

$$\text{صحت (Precision)} = \frac{\text{تعداد های نمونه تشخیصی درست مثبت}}{\text{تعداد کل های نمونه تشخیصی مثبت}} = \frac{TP}{TP+FP}$$

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

12- جدول زمانبندی مراحل انجام دادن تحقیق از زمان تصویب تا دفاع نهایی

(این جدول براساس مراحل انجام تحقیق از جمله مطالعات کتابخانه‌ای، جمع آوری اطلاعات پایه طرح، توزیع و جمع آوری پرسشنامه،مصاحبه، نمونه برداری، اندازه‌گیری، تجزیه و تحلیل داده‌ها، نگارش پایان نامه تکمیل شود)

موضوع	از تاریخ	تا تاریخ

امضا استاد راهنما

طول مدت اجرای تحقیق: ماه

تاریخ دفاع نهایی حداقل 6 ماه پس ازاخذ کد پایان نامه امکان پذیر است.

13- فهرست منابع و ماخذ (فارسی و غیر فارسی) مورد استفاده در پایان نامه:

(کلیه منابع و ماخذ بایستی در متن پروپوزال استفاده شده باشد و براساس فرمت پایین به ترتیب حروف الفبا نوشته شود):

کتاب: نام خانوادگی ، نام ، سال نشر ، عنوان کتاب ، مترجم ، محل انتشار ، جلد

مقاله: نام خانوادگی ، نام ، سال نشر ، عنوان مقاله ، عنوان نشریه ، دوره ، شماره ، صفحه

[1] O. Maimon and L. Rokach, “Data Mining and Knowledge Discovery Handbook,” pp. 1–15, 2005.

[2] H. Jang, S. Kim, and T. Lam, “Kaggle Competitions : Author Identification & Statoil / C-CORE Iceberg Classifier Challenge,” pp. 1–21, 2017.

[3] M. Kestemont et al., “Overview of the author identification task at PAN-2018: Cross-domain authorship attribution and style change detection,” CEUR Workshop Proc., vol. 2125, 2018.

[4] https://searchsqlserver.techtarget.com/definition/data-mining

[5] https://www.geeksforgeeks.org/basic-concept-classification-data-mining/

[6] https://hub.packtpub.com/what-is-ensemble-learning/

[7] https://simplicable.com/new/ensemble-learning

[8] https://www.tutorialspoint.com/data_mining/dm_knowledge_discovery.htm

[9] https://www.techopedia.com/definition/1181/data-mining

[10] https://www.geeksforgeeks.org/basic-concept-classification-data-mining/

[11] https://www.tutorialride.com/data-mining/classification-in-data-mining.htm

[12] https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html

[13] https://scikit-learn.org/stable/

[14] D. Heckerman and J. S. Breese, “Causal independence for probability assessment and inference using Bayesian networks,” IEEE Trans. Syst. Man, Cybern. Part ASystems Humans., vol. 26, no. 6, pp. 826–831, 1996.

امضا استاد راهنما

[15] A. Genkin, D.Lewis, “Author Identification on the Large Scale”, 2005

[16] https://www.edureka.co/blog/naive-bayes-tutorial/

[17] https://www.geeksforgeeks.org/understanding-logistic-regression/

[18] scikit-learn.org/stable/.../sklearn.ensemble.ExtraTreesClassifier.html

[19] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html

[20] https://blog.faradars.org/

[21] http://www.bigdata.ir/1397/03/

[22] https://howsam.org/

[23] http://research-moghimi.ir/1395/05/15/ensemble-learning/

[24] M.Brusco, D.Steinley, J.Cradit,” An exact algorithm for hierarchically well-formulated subsets in second-order polynomial regression”, 2009

[25] https://dataio.ir/how-to-import-dataset-with-pandas-r1gsggvg9xx2

[26] https://blog.faradars.org/python-excel-tutorial/

[27] https://www.datopia.ir/

[28] K. Matsumoto, “Classification of Emoji Categories from Tweet Based on Deep Neural Networks,” pp. 17–25, 2018

[29] http://www.statsoft.com/textbook/naive-bayes-classifier

[30] https://towardsdatascience.com/logistic-regression-b0af09cdb8ad

[31] https://www.quora.com/What-is-the-C-parameter-in-logistic-regression

[32] https://www.toptal.com/machine-learning/ensemble-methods-machine-learning

[33] https://xavierbourretsicotte.github.io/subset_selection.html

[34] M.Rastegar, S.Hosseinzadeh, E.Bakhshi, “Application of Logistic Regression with Missclassified Variables in Diabetes Data”, 2018

14- هزینه‌های تحقیق پایان نامه (جداول تکمیل گردد):

الف: منابع تامین بودجهٔ پایان نامه و میزان هر یک (ریالی ، ارزی ، تجهیزاتی و غیره)

ردیف	نام موسسه	بودجهٔ ریالی	بودجه ارزی	تجهیزات و تسهیلات

امضا استاد راهنما

:جمع				

ب: هزینه‌های پایان نامه

ب1: هزینه‌های پرسنلی (برای مواردی که در حوزه تخصص و مهارت و رشتهٔ دانشجو قرار ندارد)

نوع مسئولیت	تعداد افراد	کل ساعات کار برای طرح	حق‌الزحمه در ساعت	جمع
:جمع هزینه‌های تخمینی به ریال				

ب2: هزینه‌های مواد و وسایل (وسایلی که صرفاً از محل اعتبار طرح تحقیق باید خریداری شوند

نام ماده یا وسیله	مقدار مورد نیاز	مصرفی – غیر مصرفی	ساخت داخل یا خارج	شرکت سازنده	قیمت واحد		قیمت کل	
					ریالی	ارزی	ریالی	ارزی
جمع هزینه‌های مواد و وسایل به ریال								

ب3: هزینه‌های متفرقه

ردیف	شرح هزینه	ریالی	ارزی	معادل ریالی بودجهٔ ارزی	کل هزینه به ریال
1	هزینهٔ تایپ				
2	هزینهٔ تکثیر				
3	هزینهٔ صحافی				

امضا استاد راهنما

4	هزینهٔ عکس و اسلاید				
5	هزینهٔ طراحی ، خطاطی				
	نقاشی ، کارتوگرافی				
6	هزینهٔ خدمات کامپیوتری				
7	هزینه‌های دیگر				
جمع					

جمع کل هزینه‌ها

ردیف	نوع هزینه	ریالی	ارزی	هزینهٔ کل به ریال
1	پرسنلی			
2	مواد و وسایل			
3	مسافرت			
4	متفرقه			
جمع کل				

15- **تاییدات (این قسمت باید توسط اساتید تایید شود)**

الف:		
نام و نام خانوادگی استاد راهنما: عباس عکاسی	تاریخ:	امضاء
نام و نام خانوادگی مشاور:	تاریخ:	امضاء:
نام و نام خانوادگی مشاور دوم:	تاریخ:	امضاء:

ب: **نظریهٔ کمیته تخصصی گروه دربارهٔ پروپوزال**:

امضا استاد راهنما

1- ارتباط داشتن موضوع تحقیق با رشتهٔ تحصیلی دانشجو: ارتباط دارد ☐ ارتباط فرعی دارد ☐ ارتباط ندارد ☐
2- جدید بودن موضوع: بلی ☐ در ایران بلی ☐ خیر ☐
3- اهداف بنیادی و کاربردی: قابل دسترسی است ☐ قابل دسترسی نیست ☐ مطلوب نیست ☐
4- تعریف مسأله: رسا است ☐ رسا نیست ☐
5- فرضیات: درست تدوین شده است ☐ درست تدوین نشده و ناقص است ☐
6- روش تحقیق دانشجو: مناسب است ☐ مناسب نیست ☐
7- محتوا و چهارچوب طرح: از انسجام برخوردار است ☐ از انسجام برخوردار نیست ☐

امضا استاد راهنما

موضوع پایان نامهٔ خانم / آقای:
دانشجوی مقطع: کارشناسی ارشد■ دکترای حرفه‌ای□ رشته:
تحت عنوان:
در جلسه مورخ کمیتهٔ تخصصی گروه مطرح شده و به اتفاق آرا با تعداد... رای از.... رای مورد تصویب اعضاء قرار گرفت □ قرار نگرفت □

مدیر گروه تاریخ: امضاء

پ: تائید نهایی- شورای تخصصی گروه

ردیف	نام و نام خانوادگی	سمت و تخصص	نوع رأی	امضاء
1				
2				
3				
4				
5				

ث: نظریهٔ شورای پژوهشی دانشکده:
موضوع و طرح تحقیق پایان نامهٔ آقای / خانم دانشجوی مقطع:
رشتهٔ که به تصویب کمیته تخصصی مربوط رسیده بود در جلسه مورخ:
شورای پژوهشی دانشگاه مطرح شد و پس از بحث و تبادل نظر مورد تصویب اکثریت اعضاء (تعداد نفر)
قرار گرفت / نگرفت .

ردیف	نام و نام خانوادگی	نوع رأی (موافق یا مخالف)	امضاء	توضیحات
1				
2				
3				
4				
5				
6				
7				

نام و نام خانوادگی معاون پژوهشی واحد	تاریخ	امضاء
شمارۀ ثبت در امور پژوهشی واحد	تاریخ ثبت	