



دانشگاه اصفهان

دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش
هوش مصنوعی

کلاس بندی هوشمند ترافیک شبکه بر مبنای مشخصات محتوی جریان و
اطلاعات آماری ترافیک

استادان راهنما:

دکتر ناصر موحدی نیا

دکتر محمدرضا خیام باشی

پژوهشگر:

فرشته دهقانی

آبان ماه ۱۳۸۹

کلیه حقوق مادی مترتب بر نتایج مطالعات،
ابتکارات و نوآوری های ناشی از تحقیق موضوع
این پایان نامه متعلق به دانشگاه اصفهان است.



دانشگاه اصفهان

دانشکده فنی مهندسی

گروه مهندسی کامپیوتر

پایان نامه کارشناسی ارشد رشته مهندسی کامپیوتر گرایش

هوش مصنوعی خانم فرشته دهقانی تحت عنوان

کلاس بندی هوشمند ترافیک شبکه بر مبنای مشخصات محتوی جریان و اطلاعات

آماري ترافیک

در تاریخ ۱۳۸۹/۰۸/۰۳ توسط هیأت داوران زیر بررسی و با درجه عالی به تصویب نهایی رسید.

امضا

۱- استاد راهنمای اول پایان نامه دکتر ناصر موحدی نیا با مرتبه علمی دانشیار

امضا

۲- استاد راهنمای دوم پایان نامه دکتر محمدرضا خیامباشی با مرتبه علمی استادیار

امضا

۳- استاد داور داخل گروه دکتر محمد علی نعمت بخش با مرتبه علمی دانشیار

امضا

۴- استاد داور خارج از گروه دکتر حسین سعیدی با مرتبه علمی دانشیار

امضای مدیر گروه

امضا

سپاس بی دریغ خود را نثار استاد با فضیلت جناب آقای دکتر ناصر موحدی نیامی نمایم که در کلیه مراحل تحصیل، با فضل و دانش خود راهنمایی

ایجناب بوده اند.

از جناب آقای دکتر محمد رضا خیام باشی که ناهمواریها و تاریکی های این راه صعب را با پرتو نور دانش خود، هموار و روشن نمودند شکر و قدر دانی می

نمایم.

به پاس تعبیر عظیم و انسانی شان از کلمه ایثار و از خودگذشتگی

به پاس عاطفه سرشار و گرمای امید بخش وجودشان که در این سردترین روزگار ان بهترین پشتیبان است

به پاس قلب های بزرگشان که فریاد رس است و سرگردانی و ترس در پناهشان به شجاعت می گراید

و به پاس محبت های بی دریغشان که هرگز فروکش نمی کند

این مجموعه را به خانواده عزیزم تقدیم می کنم.

چکیده

در شبکه‌های امروزی، کاربردهای بسیار متنوعی ارائه می‌شوند که سرویس‌های متفاوت با نیازمندی‌های ترافیکی خاص خود را طلب می‌کنند. به این جهت در مدیریت شبکه، تشخیص و کلاس‌بندی ترافیک نقش بسیار مهمی در بالا بردن بهره‌وری شبکه بازی می‌کند.

در روش‌های اولیه کلاس‌بندی ترافیک، نوع ترافیک با توجه به شماره پورت تشخیص داده می‌شد. اما با افزایش کاربردهایی که از شماره پورت‌های ناشناخته استفاده می‌کردند، کارایی این روش‌ها کاهش یافت. روش‌های بعدی بر مبنای بازرسی عمقی بسته‌ها بودند. این روش‌ها علیرغم صحت عملکرد مناسب، نیازمند دسترسی به محتوای پیام بودند که به دلیل مسائل امنیتی همیشه امکان‌پذیر نیست و همچنین پیچیدگی محاسباتی بالایی نیاز دارند. روش‌های متاخر با استفاده از مشخصات آماری بسته‌ها و الگوریتم‌های یادگیری ماشین، ترافیک را دسته‌بندی می‌کنند. اما در کلاس‌بندی بلادرنگ، معمولاً تنها به بخش کمی از اطلاعات آماری جریان می‌توان دسترسی داشت.

در این رساله با توجه به مزایا و معایب دو روش اخیر، ترافیک شبکه بر مبنای اطلاعات محتوای جریان و مشخصات آماری ترافیک، با استفاده از الگوریتم‌های بی‌زین و مبتنی بر درخت تصمیم کلاس‌بندی شده است. به این منظور از الگوریتم‌های انتخاب ویژگی و نمونه‌برداری برای پیش‌پردازش مجموعه داده استفاده شده و از کواریانس ویژگی‌ها برای کشف وابستگی میان بسته‌های هر جریان، بهره‌برده می‌شود. نتایج نشان می‌دهند که الگوریتم‌های مبتنی بر درخت و کواریانس ویژگی‌ها به بهترین جواب‌های کلاس‌بندی می‌رسند، در حالی که الگوریتم بی‌زین ساخت مدل سریع‌تری نسبت به الگوریتم‌های مبتنی بر درخت دارد.

همچنین به منظور کلاس‌بندی بلادرنگ ترافیک شبکه، از زیرجریان‌ها و ماتریس کواریانس بهره‌برده شده است. نتایج نشان می‌دهد که کواریانس زیرجریان‌ها صحت و سرعت بهتری را به نسبت استفاده از تمام بسته‌های هر جریان دارد. به گونه‌ای که تنها با استفاده از شانزده بسته‌ی ابتدای جریان، صحت کلاس‌بندی بالاتر از ۰/۹۹۵ برای تمام کاربردها به دست می‌آید. این امر مناسب بودن روش پیشنهادی را برای کلاس‌بندی بلادرنگ ترافیک تایید می‌نماید.

واژگان کلیدی: خصوصیات آماری ترافیک شبکه، کلاس‌بندی هوشمند ترافیک شبکه، محتوای جریان، یادگیری ماشین.

فهرست مطالب

صفحه

عنوان

فصل اول: پیشگفتار

فصل دوم: زمینه تحقیق

۷	۱-۲ مقدمه
۸	۲-۲ معماری TCP/IP
۹	۱-۲-۲ لایه انتقال
۱۰	۲-۲-۲ مدل ارتباطی TCP
۱۳	۳-۲-۲ مدل ارتباطی UDP
۱۵	۳-۲ کلاس بندی بر مبنای پورت
۱۸	۴-۲ کلاس بندی بر مبنای محتوای بسته ها
۱۸	۱-۴-۲ روش های واریسی
۱۹	۲-۴-۲ کلاس بندی کاربرد
۲۱	۳-۴-۲ روش های پردازش مبتنی بر محتوا
۲۲	۵-۲ یادگیری ماشین و کاربرد آن در کلاس بندی ترافیک شبکه
۲۳	۱-۵-۲ یادگیری ماشین در کلاس بندی ترافیک
۲۴	۲-۵-۲ یادگیری با نظارت
۲۶	۳-۵-۲ یادگیری بدون نظارت
۲۸	۶-۲ جمع بندی

فصل سوم: مروری بر پژوهش های پیشین

۳۰	۱-۳ مقدمه
۳۱	۲-۳ طرح موضوع
۳۲	۳-۳ روش های یادگیری بدون نظارت
۳۳	۱-۳-۳ خوشه بندی جریان با استفاده از الگوریتم EM
۳۳	۲-۳-۳ خوشه بندی ترافیک با استفاده از AutoClass
۳۴	۳-۳-۳ خوشه بندی ترافیک در حال تغییر
۳۴	۴-۳-۳ تفکیک و تمایز ترافیک وب و P2P

۳-۳-۵ کلاس بندی ترافیک با استفاده از الگوریتم های بدون نظارت و KNN.....	۳۵
۳-۴-۴ روش های یادگیری بانظارت	۳۵
۳-۴-۱ روش مبتنی بر امضای آماری با استفاده از روش های یادگیری ماشین.....	۳۶
۳-۴-۲ کلاس بندی ترافیک با استفاده از الگوریتم و شبکه های بیزین.....	۳۶
۳-۴-۳ کلاس بندی ترافیک با استفاده از زیرجریان های چندگانه.....	۳۷
۳-۴-۴ کلاس بندی ترافیک با استفاده از آنالیز تجربی سطح کاربرد	۳۸
۳-۴-۵ کلاس بندی ترافیک به صورت بلادرنگ	۳۸
۳-۴-۶ الگوریتم های یادگیری ماشین به منظور کلاس بندی شبکه	۳۹
۳-۵-۵ روش های نیمه نظارتی	۳۹
۳-۵-۱ کلاس بندی نیمه نظارتی شبکه با استفاده از KMeans.....	۳۹
۳-۵-۲ کلاس بندی نیمه نظارتی شبکه با استفاده از مدل ترکیبی گاوسین.....	۴۰
۳-۶-۶ استفاده از محتوای داده به عنوان ویژگی	۴۱
۳-۶-۱ ساخت اتوماتیک امضاهای کاربرد با استفاده از یادگیری ماشین.....	۴۱
۳-۶-۲ روش بدون نظارت به منظور استنباط کاربرد با استفاده از محتوای جریان.....	۴۲
۳-۶-۳ تشخیص پروتکل لایه کاربرد با استفاده از اندازه محتوای پیام.....	۴۳
۳-۷-۷ روش های دیگر کلاس بندی	۴۳
۳-۷-۱ تشخیص حمله شبکه در فضای کواریانس ویژگی.....	۴۳
۳-۷-۲ کلاس بندی ترافیک با استفاده از جداسازی کور منبع.....	۴۴
۳-۸ جمع بندی	۴۴

فصل چهارم: رهیافت پیشنهادی به منظور کلاس بندی ترافیک شبکه

۴-۱ مقدمه.....	۴۶
۴-۲ مجموعه داده	۴۸
۴-۲-۱ مشخصات مجموعه داده.....	۴۸
۴-۲-۲ استخراج جریان ها از بسته ها.....	۵۳
۴-۳ محاسبه آمار ترافیک	۵۵
۴-۳-۱ ویژگی های آماری جریان بسته ها.....	۵۵

۲-۳-۴ ویژگی های جریان با استفاده از مشخصات آماری و محتوای بسته ها	۵۶
۴-۴ پیش پردازش داده	۵۸
۱-۴-۴ الگوریتم پیش پردازش نمونه ها (نمونه برداری)	۵۸
۲-۴-۴ الگوریتم پیش پردازش ویژگی ها (انتخاب ویژگی)	۵۹
۵-۴ استفاده از کواریانس ویژگی ها	۶۲
۶-۴ الگوریتم یادگیری بانظارت	۶۴
۱-۶-۴ یادگیری بیزین	۶۴
۲-۶-۴ الگوریتم یادگیری ماشین مبتنی بر درخت	۶۷
۷-۴ کلاس بندی بلادرنگ	۷۴
۸-۴ پیشنهاد مدلی برای کلاس بندی به صورت برخط	۷۶
۹-۴ جمع بندی	۷۷

فصل پنجم: شبیه سازی و ارزیابی نتایج

۱-۵ مقدمه	۷۸
۲-۵ معیارهای ارزیابی کلاس بندی ترافیک شبکه	۷۹
۳-۵ نتایج کلاس بندی ترافیک شبکه	۸۱
۱-۳-۵ نتایج برگزیده الگوریتم بیزین	۸۵
۲-۳-۵ نتایج برگزیده الگوریتم C4.5	۸۷
۳-۳-۵ نتایج برگزیده الگوریتم RandomForest	۹۰
۴-۳-۵ نتایج برگزیده الگوریتم C5.0	۹۳
۵-۳-۵ مقایسه نتایج	۹۴
۴-۵ نتایج کلاس بندی بلادرنگ ترافیک با استفاده از کواریانس ویژگی و زیرجریان	۹۵
۵-۵ مقایسه با روش کلاس بندی ترافیک به صورت بلادرنگ	۹۶
۶-۵ مقایسه با روش تشخیص پروتکل لایه کاربرد با استفاده از اندازه محتوای پیام	۹۸
۷-۵ جمع بندی	۹۹

فصل ششم: جمع بندی و نتیجه گیری

پیوست ها

پیوست الف.....	۱۰۵
الف-۱ توضیح در مورد نرم افزار Weka.....	۱۰۵
الف-۲ توضیح در مورد نرم افزار SPSS Clementine.....	۱۰۷
پیوست ب.....	۱۰۹
ب-۱- نتایج کلاس بندی با استفاده از مجموعه داده بدون پیش پردازش.....	۱۰۹
ب-۲- نتایج کلاس بندی با استفاده از مجموعه داده و پیش پردازش نمونه ها.....	۱۱۶
ب-۳- نتایج کلاس بندی با استفاده از مجموعه داده و پیش پردازش ویژگی.....	۱۲۹
ب-۴- نتایج کلاس بندی با استفاده از مجموعه داده و پیش پردازش نمونه ها و ویژگی.....	۱۳۵
پیوست پ.....	۱۵۳
پ-۱ کلاس بندی بلادرنگ با استفاده از زیرجریان و کواریانس ویژگی.....	۱۵۳
منابع و مآخذ:	۱۵۸

فهرست شکل ها

عنوان	صفحه
شکل ۱-۲- TCP/IP مدل	۸
شکل ۲-۲- ارتباط منطقی در میزبان سرویس گیرنده و سرویس دهنده در لایه انتقال	۹
شکل ۳-۲- پروتکل های TCP و UDP در ارتباط با پروتکل اینترنت	۱۰
شکل ۴-۲- ساختار بسته TCP	۱۱
شکل ۵-۲- مدل PAR در یک عملیات عادی	۱۲
شکل ۶-۲- پنجره متحرک TCP به منظور کنترل جریان	۱۲
شکل ۷-۲- ساختار سرآیند UDP	۱۴
شکل ۸-۲- روش های بازرسی و پردازشی متفاوت در روش بازرسی بر مبنای محتوای بسته ها	۱۹
شکل ۹-۲- دامنه کاربرد در مقابل دامنه پروتکل	۲۰
شکل ۱۰-۲- روند الگوریتم های یادگیری با نظارت	۲۶
شکل ۱۱-۲- روند الگوریتم های یادگیری بدون نظارت	۲۷
شکل ۱-۴- روند کلاس بندی ترافیک شبکه با استفاده از الگوریتم یادگیری ماشین	۴۷
شکل ۲-۴- نمودار مجموعه داده استفاده شده	۵۲
شکل ۳-۴- شرح کلاس بندی در فضای کواریانس	۶۳
شکل ۴-۴- الگوریتم اولیه برای استنتاج یک درخت تصمیم از نمونه های آموزش	۶۹
شکل ۵-۴- سه نوع بخشبندی ممکن نمونه ها بر مبنای معیار جداسازی در درخت تصمیم	۷۱
شکل ۱-۵- طبقه بندی مجموعه داده و الگوریتم های استفاده شده	۸۴
شکل ۲-۵- نتایج الگوریتم بیزین با استفاده از مجموعه داده بدون پیش پردازش	۸۵
شکل ۳-۵- نتایج الگوریتم بیزین با استفاده از الگوریتم های نمونه برداری	۸۶
شکل ۴-۵- نتایج الگوریتم بیزین با استفاده از الگوریتم های انتخاب ویژگی	۸۷
شکل ۵-۵- نتایج الگوریتم بیزین با استفاده از الگوریتم های انتخاب ویژگی و نمونه برداری	۸۷
شکل ۶-۵- نتایج الگوریتم C4.5 با استفاده از مجموعه داده بدون پیش پردازش	۸۸
شکل ۷-۵- نتایج الگوریتم C4.5 با استفاده از الگوریتم های نمونه برداری	۸۹
شکل ۸-۵- نتایج الگوریتم C4.5 با استفاده از الگوریتم های انتخاب ویژگی	۸۹
شکل ۹-۵- نتایج الگوریتم C4.5 با استفاده از الگوریتم های انتخاب ویژگی و نمونه برداری	۹۰

- شکل ۵-۱۰- نتایج الگوریتم RandomForest با استفاده از مجموعه داده بدون پیش پردازش ۹۱
- شکل ۵-۱۱- نتایج الگوریتم RandomForest با استفاده از الگوریتم های نمونه برداری ۹۱
- شکل ۵-۱۲- نتایج الگوریتم RandomForest با استفاده از الگوریتم های انتخاب ویژگی ۹۲
- شکل ۵-۱۳- نتایج الگوریتم RandomForest با استفاده از الگوریتم های انتخاب ویژگی و نمونه برداری ۹۳
- شکل ۵-۱۴- نتایج الگوریتم C5.0 با استفاده از مجموعه داده بدون پیش پردازش ۹۳
- شکل ۵-۱۵- نتایج الگوریتم C5.0 با استفاده از الگوریتم های انتخاب ویژگی و نمونه برداری ۹۴
- شکل ۵-۱۶- مقایسه نتایج الگوریتم های استفاده شده ۹۵
- شکل ۵-۱۷- نتایج کلاس بندی بلادرنگ با استفاده از کواریانس ویژگی و زیرجریان ها ۹۶
- شکل ۵-۱۸- مقایسه نتایج کلاس بندی ترافیک به صورت بلادرنگ و رهیافت پیشنهادی با استفاده تمام جریان ۹۷
- شکل ۵-۱۹- مقایسه نتایج روش تشخیص پروتکل لایه کاربرد با استفاده از اندازه محتوای پیام و رهیافت پیشنهادی ۹۸
- شکل الف-۱- تعامل سرویس دهنده و سرویس گیرنده در Clementine ۱۰۸
- شکل ب-۱- از سمت راست ۸۰ درصد داده آموزش، ۶۶ درصد داده آموزش، واری اعتبار $K=6$ ، $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده - ۶۶درصد داده آموزش با استفاده از الگوریتم C4.5 ۱۰۹
- شکل ب-۲- ۶۶درصد داده آموزش با استفاده از الگوریتم C5.0 ۱۱۰
- شکل ب-۳- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده - ۶۶ درصد داده آموزش، واری اعتبار $K=10$ با استفاده از الگوریتم C4.5 ۱۱۰
- شکل ب-۴- ۶۶ درصد داده آموزش با استفاده از الگوریتم C5.0 ۱۱۱
- شکل ب-۵- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده - ۶۶ درصد داده آموزش، واری اعتبار $K=10$ با استفاده از الگوریتم C4.5 ۱۱۱
- شکل ب-۶- ۶۶درصد داده آموزش با استفاده از الگوریتم C5.0 ۱۱۲
- شکل ب-۷- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده - ۶۶ درصد داده آموزش، واری اعتبار $K=10$ با استفاده از الگوریتم C4.5 ۱۱۲
- شکل ب-۸- ۶۶درصد داده آموزش با استفاده از الگوریتم C5.0 ۱۱۳

شکل ب-۹- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده- ۶۶ درصد داده آموزش، واری اعتبار $K=10$ با استفاده از الگوریتم C4.5 ۱۱۳

شکل ب-۱۰- ۶۶درصد داده آموزش با استفاده از الگوریتم C5.0..... ۱۱۴

شکل ب-۱۱- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده- ۶۶ درصد داده آموزش، واری اعتبار $K=10$ با استفاده از الگوریتم C4.5 ۱۱۴

شکل ب-۱۲- ۶۶درصد داده آموزش با استفاده از الگوریتم C5.0..... ۱۱۵

شکل ب-۱۳- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده- ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 - ۶۶ درصد داده آموزش با استفاده از الگوریتم RandomForest..... ۱۱۵

شکل ب-۱۴- ۶۶درصد داده آموزش با استفاده از الگوریتم C5.0..... ۱۱۶

شکل ب-۱۵- از راست به چپ ۶۶ درصد داده آموزش، ۸۰ درصد داده آموزش، واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ۱۱۶

شکل ب-۱۶- از راست به چپ واری اعتبار $K=8$ و $K=10$ ، ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقدار $S=2$ ۱۱۷

شکل ب-۱۷- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ، ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 و الگوریتم SpreadSubSample و مقدار $S=2$ ۱۱۷

شکل ب-۱۸- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم RandomForest و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ، ۶۶ درصد داده آموزش با استفاده از الگوریتم RandomForest و الگوریتم SpreadSubSample و مقدار $S=2$ ۱۱۸

شکل ب-۱۹- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ، با مقادیر $B=1$ و $P=100$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقدار $S=0$ ، $S=2$ ۱۱۸

شکل ب-۲۰- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های C4.5 و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم های RandomForest و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ ۱۱۹

شکل ب-۲۱- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ، با مقادیر $B=1$ و $P=100$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقدار $S=0$ ، $S=2$ ۱۲۰

شکل ب-۲۲- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های C4.5 و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم های RandomForest و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ ۱۲۰

شکل ب-۲۳- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقادیر $B=1$ و $P=80$ - $B=1$ و $P=100$ ۱۲۱

شکل ب-۲۴- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقدار $S=0$ ، $S=2$ ۱۲۱

شکل ب-۲۵- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های C4.5 و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم های RandomForest و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ ۱۲۲

شکل ب-۲۶- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقادیر $B=1$ و $P=80$ - $B=1$ و $P=100$ ۱۲۳

شکل ب-۲۷- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقدار $S=0$ ، $S=2$ ۱۲۳

- شکل ب-۲۸- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های C4.5 و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم های RandomForest و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ ۱۲۴
- شکل ب-۲۹- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ، $B=1$ و $P=100$ - واری اعتبار $K=8$ ، $K=10$ و الگوریتم Resample با مقادیر $B=1$ و $P=80$ ۱۲۴
- شکل ب-۳۰- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقدار $S=2$ ، $S=0$ ۱۲۵
- شکل ب-۳۱- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های C4.5 و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ - ۶۶ درصد داده آموزش با استفاده از الگوریتم های RandomForest و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ ۱۲۵
- شکل ب-۳۲- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم Resample با مقدار $P=80$ ، $B=1$ و $P=100$ ، $B=1$ و $P=80$ ، $B=1$ و $K=10$ اعتبار $P=80$ ، $B=1$ ۱۲۶
- شکل ب-۳۳- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم بیزین ساده و الگوریتم SpreadSubSample با مقادیر $S=2$ ، $S=0$ ، واری اعتبار $K=8$ ، $K=10$ و $S=2$ ۱۲۷
- شکل ب-۳۴- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های C4.5 و Resample با مقادیر $B=1$ و $P=80$ ، الگوریتم SpreadSubSample با مقدار $S=2$ ۱۲۷
- شکل ب-۳۵- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم های RandomForest و Resample با مقادیر $B=1$ و $P=80$ ، $B=1$ و $P=100$ - الگوریتم SpreadSubSample با مقدار $S=2$ ، $S=0$ ۱۲۸
- شکل ب-۳۶- از راست به چپ ۶۶ درصد آموزش و ۸۰ درصد آموزش با استفاده از الگوریتم بیزین و جستجوی اول بهترین - جستجوی ژنتیک ۱۲۹
- شکل ب-۳۷- از راست به چپ واری اعتبار $K=8$ و $K=10$ با استفاده از الگوریتم بیزین و جستجوی اول بهترین - جستجوی ژنتیک ۱۳۰

- شکل ب-۳۸- از راست به چپ ۶۶ درصد آموزش با استفاده از الگوریتم C4.5 و جستجوی اول بهترین -
 جستجوی ژنتیک ۱۳۰
- شکل ب-۳۹- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم بیزین و
 جستجوی اول بهترین، جستجوی ژنتیک ۱۳۱
- شکل ب-۴۰- از راست به چپ ۶۶ درصد آموزش با استفاده از الگوریتم C4.5 و جستجوی اول بهترین -
 جستجوی ژنتیک ۱۳۱
- شکل ب-۴۱- از راست به چپ ۶۶ درصد آموزش با استفاده از الگوریتم Random Forest و جستجوی اول
 بهترین - جستجوی ژنتیک ۱۳۲
- شکل ب-۴۲- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم بیزین و
 جستجوی ژنتیک ۱۳۲
- شکل ب-۴۳- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 و جستجوی اول بهترین -
 با استفاده از الگوریتم Random Forest ۱۳۲
- شکل ب-۴۴- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم های
 بیزین ساده و جستجوی اول بهترین ۱۳۳
- شکل ب-۴۵- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 و جستجوی اول بهترین -
 با استفاده از الگوریتم Random Forest ۱۳۳
- شکل ب-۴۶- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم های
 بیزین ساده و جستجوی اول بهترین ۱۳۳
- شکل ب-۴۷- از راست به چپ ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 و جستجوی اول بهترین -
 با استفاده از الگوریتم Random Forest ۱۳۴
- شکل ب-۴۸- از راست به چپ ۶۶ درصد داده آموزش و واریسی اعتبار $K=10$ با استفاده از الگوریتم بیزین
 جستجوی ژنتیک ۱۳۴
- شکل ب-۴۹- از راست به چپ واریسی ۶۶ درصد داده آموزش با استفاده از الگوریتم C4.5 و جستجوی اول
 بهترین ۱۳۴
- شکل ب-۵۰- از راست به چپ ۶۶ درصد داده های آموزش با استفاده از الگوریتم های بیزین و Resample،
 اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک ۱۳۵

- شکل ب-۵۱- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۶
- شکل ب-۵۲- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های RandomForest و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۶
- شکل ب-۵۳- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۷
- شکل ب-۵۴- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۸
- شکل ب-۵۵- از راست به چپ واریسی اعتبار $K=10$ با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۸
- شکل ب-۵۶- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۹
- شکل ب-۵۷- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های Random Forest و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۳۹
- شکل ب-۵۸- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۰
- شکل ب-۵۹- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۰
- شکل ب-۶۰- از راست به چپ واریسی اعتبار $K=10$ با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۱
- شکل ب-۶۱- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۱
- شکل ب-۶۲- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های RandomForest و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۲
- شکل ب-۶۳- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۲

- شکل ب-۶۴- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۳
- شکل ب-۶۵- از راست به چپ واریسی اعتبار $K=10$ با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۴
- شکل ب-۶۶- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۴
- شکل ب-۶۷- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های RandomForest و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۵
- شکل ب-۶۸- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۵
- شکل ب-۶۹- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های بیزین و Resample، اول بهترین - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۶
- شکل ب-۷۰- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample، اول بهترین - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۶
- شکل ب-۷۱- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های RandomForest و Resample، اول بهترین - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۷
- شکل ب-۷۲- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample، اول بهترین - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۷
- شکل ب-۷۳- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های بیزین و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۸
- شکل ب-۷۴- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۹
- شکل ب-۷۵- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های RandomForest و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۴۹
- شکل ب-۷۶- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample، اول بهترین و ژنتیک - الگوریتم SpreadSubSample، اول بهترین و ژنتیک..... ۱۵۰

شکل ب-۷۷- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های بیزین و Resample.	
اول بهترین و ژنتیک - الگوریتم SpreadSubSample و ژنتیک	۱۵۰
شکل ب-۷۸- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C4.5 و Resample.	
اول بهترین و ژنتیک - الگوریتم SpreadSubSample و ژنتیک	۱۵۱
شکل ب-۷۹- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های RandomForest و Resample.	
اول بهترین و ژنتیک - الگوریتم SpreadSubSample و ژنتیک	۱۵۱
شکل ب-۸۰- از راست به چپ ۶۶ درصد داده‌های آموزش با استفاده از الگوریتم های C5.0 و Resample.	
اول بهترین و ژنتیک - الگوریتم SpreadSubSample و ژنتیک	۱۵۲
شکل پ-۱- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های بیزین و Resample.	
.....	۱۵۳
شکل پ-۲- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های RandomForest و Resample.	
.....	۱۵۴
شکل پ-۳- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های C4.5 و Resample.	
.....	۱۵۴
شکل پ-۴- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های C5.0 و Resample.	
.....	۱۵۵
شکل پ-۵- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های بیزین و Resample.	
.....	۱۵۵
شکل پ-۶- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های C4.5 و Resample.	
.....	۱۵۶
شکل پ-۷- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های RandomForest و Resample.	
.....	۱۵۶
شکل پ-۸- از راست به چپ ۴، ۸، ۱۶ و ۳۲ بسته ابتدایی جریان با استفاده از الگوریتم های C5.0 و Resample.	
.....	۱۵۷

فهرست جدول ها

صفحه	عنوان
۱۶	جدول ۱-۲- شماره پورتهای موجود در IANA برای تعدادی از کاربردها و پروتکل ها
۵۰	جدول ۱-۴- مشخصات مجموعه داده ی اول
۵۱	جدول ۲-۴- مشخصات مجموعه داده ی دوم
۵۱	جدول ۳-۴- مشخصات مجموعه داده ی سوم
۵۲	جدول ۴-۴- مشخصات کلی مجموعه داده
۸۰	جدول ۱-۵- رابطه میان ملاک های ارزیابی
۸۲	جدول ۲-۵- توضیح پارامترهای الگوریتم Resample
۸۲	جدول ۳-۵- توضیح پارامترهای الگوریتم SpreadSubSample