بررسی موردی انتخاب خصیصه‌های مناسب برداشت سیب با استفاده از یادگیری ماشین (۲۴۷)

آرش معبودی[۱] ، احسان کشتکاری[۲] ، محمد امین فرجیان[۳]

چکیده

در عرصه کشاورزی، تعیین زمان برداشت محصول اهمیت ویژه‌ای دارد، که برای این کار از فاکتورهای مختلفی استفاده می شود. در این مقاله به بررسی موردی انتخاب خصیصه مناسب برای تشخیص وضعیت سیب و تعیین زمان برداشت محصول آن با استفاده از تکنیک های یادگیری ماشین پرداخته شده است. با کمک افراد خبره، کلاس های لازم برای محصول سیب، خصیصه های مهم در تعیین وضعیت سیب و دامنه ارزش خصیصه‌ها تعیین شده و به همین ترتیب یک مجموعه داده از ۱۰۰ نمونه سیب تهیه شده‌است. در ادامه الگوریتم های آموزشی مختلف را روی مجموعه داده آزموده و با کلاس بندی نمونه‌ها بهترین خصیصه سیب برای تعیین زمان برداشت مشخص می‌شود. نوع پراکندگی نمونه‌ها به ازای هر خصیصه، مشخص کننده بهترین خصیصه سیب از میان ۵ خصیصه خواهد بود. برای اعمال الگوریتم‌های آموزشی از نرم‌افزار وکا استفاده شده است.

۱- دانشجوی کارشناسی ارشد، هوش مصنوعی، دانشگاه علم و صنعت ایران پست الکترونیک: maboodi@comp.iust.ac.ir
۲- دانشجوی کارشناسی ارشد، هوش مصنوعی، دانشگاه علم و صنعت ایران
۳- دانشجوی کارشناسی ارشد، هوش مصنوعی، دانشگاه علم و صنعت ایران

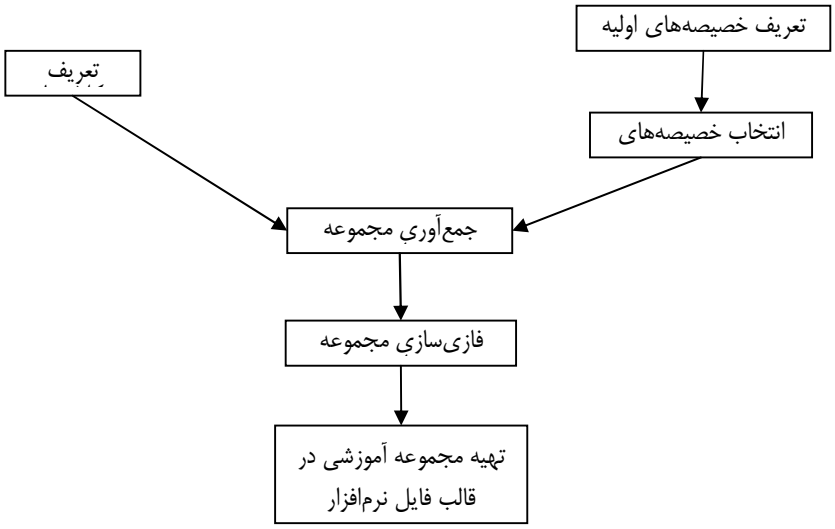
مقدمه

یادگیری ماشین[1] یکی از مهمترین شاخه های تحقیقاتی هوش مصنوعی است. هدف از تحقیقات در این زمینه، دستیابی به فنونی است که امکان شبیه سازی رفتار یادگیری[2] به عنوان یک رفتار هوشمندانه را به وسیله رایانه فراهم کنند. هدف محققین، تولید ماشینهایی است که قابلیت یادگیری دارند زیرا یادگیری، یک رفتار مهم انسان محسوب می شود. یادگیری سبب می شود که انسان دانش خود را نسبت به محیط اطراف افزایش دهد. در ماشین یادگیری وقتی رخ می‌دهد که ماشین بتواند برنامه، ساختار یا داده‌هایش را براساس ورودی‌ها یا در پاسخ به اطلاعات خارجی به نحوی تغییر دهد که رفتارش به آن‌چه از او انتظار می‌رود، نزدیک‌تر شود.

تعریف مساله

هدف از این پروژه مشاهده و بررسی سیب به عنوان یک محصول کشاورزی و تصمیم‌گیری درخصوص وضعیت آن می‌باشد.

روند تعریف این پروژه بدین صورت بوده است که ابتدا کلاسهای مورد نظر در خصوص این محصول تعیین می‌شوند. پس از تعیین کلاسها، خصیصه‌های[3] مهم در تعیین وضعیت سیب توسط خبره و کسانی که در مورد خصوصیات سیب شناخت کافی دارند، تعیین شده و دامنه ارزش[4] خصیصه‌ها نیز توسط خبرگان تعیین می‌شود. سپس مرحله بررسی و انتخاب خصیصه‌های[5] برتر با مشورت خبره انجام گرفته و نهایی می‌شود. در پایان نیز مجموعه داده[6] یا مجموعه آموزشی[7] مورد نظر براساس کلاسها و خصیصه‌های انتخاب شده به صورت داده‌های خام بر اساس مشاهدات موجود جمع‌آوری شده‌اند. شکل ۱ نمایانگر مراحل انجام روند فعالیت تعریف مسله می‌باشد.



شکل۱: مراحل انجام روند فعالیت تعریف مساله

1 Machin Learning
2 Learning
3 Attribute
4 Value
5 Attribute selection
6 Data Set
7 Traning Set

تعریف کلاسها

در بررسی و تعیین کلاس سیب به عنوان محصول مورد نظر، نکته قابل توجه این است که با توجه ماهیت این محصول در منطقه رنگ غالب آن دو رنگ قرمز و طلایی می‌باشد که با در نظر گرفتن دو حالت نارس و رسیده برای سیب چهار کلاس مختلف برای این مساله قابل تعریف می‌باشد. این چهار کلاس عبارتند از:

Ripe red(RR)قرمز رسیده
Ripe Golden(RG)طلایی رسیده
Unripe red(UR)قرمز نارس
Unripe Golden(UG)طلایی نارس
توجه: رنگهای کلاسها در شکلها به صورت بالا است:

تعریف خصیصه‌ها

در مرحله بعد پس از مصاحبه با چندین خبره در خصوص خصوصیات سیب خصیصه‌های اولیه زیر تعیین شد:

۱. **تعداد روزهای تمام گل:** مدت زمان طی شده از روزهای تمام گل شده درخت را گویند. (روز تمام گل روزی است که حداقل ۹۰درصد گلهای درخت شکوفه زده باشد.)
۲. **رنگ پوست میوه:** منظور رنگ غالب زمینه میوه می‌باشد که به راحتی قابل مشاهده است.
۳. **رنگ دون میوه:** رنگ دون یا هسته میوه می‌باشد.
۴. **میزان سفتی میوه:** این مقدار با یک دستگاه فشارسنج مخصوص اندازه گیری می‌شود و میزان فرورفتگی به وسیله دستگاه فشارسنج به میلی‌متر نشانگر سفتی میوه است.
۵. **درصد شناسته:** برای تعیین این مقدار نیاز به کار آزمایشگاهی دارد، بدین صورت که میوه بریده می‌شود و محلولی خاص روی میوه ریخته شده و میزان تغییر رنگ میوه نشانگر درصد نشاسته آن است.
۶. **وضعیت محیطی منطقه:** اعم از وضعیت خاک، استعداد و باروری محیط
۷. **آب و هوای محیط:** با توجه به تغییرات آب و هوا در هر سال و احتمال خشکسالی یا بارندگی زیاد برای یک محیط ثابت، آب و هوای آن سال به عنوان یک خصیصه مطرح می‌شود.

انتخاب خصیصه‌ها

هدف از این مرحله انتخاب خصیصه‌هایی است که با هدف ما بیشتر منطبق باشند. در این مرحله از نظر خبره در انتخاب خصیصه-های برتر استفاده می‌کنیم. در پروژه فعلی با مشورت و دریافت نظر خبرگی ۵ خصیصه زیر به عنوان خصیصه‌های برتر و مهمتر تعیین شدند. این خصیصه‌ها عبارتند از: ۱- تعداد روزهای تمام گل ۲- رنگ پوست میوه ۳- رنگ دون میوه ۴- میزان سفتی میوه ۵- درصد شناسته

نکته مهم:

لازم به توضیح است که مشاهده خصیصه ۳،۴و۵ مجموعه داده را دچار آسیب کرده و میوه‌های مجموعه داده مشاهده شده قابل استفاده نمی‌باشد.

تعیین محدوده مقادیر خصیصه‌ها

در این مرحله با دریافت نظر از خبره محدوده معقول مقادیر خصیصه‌ها را معین می‌کنیم. این محدوده مقادیر بر اساس تجربیات قبلی در خصوص مقادیر قابل رویت هر خصیصه تعیین می‌شود.

۱. تعداد روزهای تمام گل (Days)
   - بین ۱۳۰-۱۳۶روز (برای سیب قرمز)
   - بین ۱۳۶-۱۴۰ روز (برای سیب طلایی)

۲. رنگ پوست میوه (Fruit-Color)
- سبز روشن lightgreen
- زرد مایل به سبز yellowgreen
- سبز با رگه های قرمز redgreen
- قرمز red

۳. رنگ دون میوه (Kernel-Color)
- قهوه‌ای brown
- سفید مایل به زرد yellow

۴. میزان سفتی میوه (Fruit-Tenacity)
- از ۵ تا ۱۲ میلیمتر

۵. درصد شناسته (Fruit-Percent)
- کم
- زیاد

جمع‌آوری مجموعه داده

در این مرحله هدف جمع‌آوری مجموعه داده بر اساس مشاهدات و تعیین مقادیر خصیصه‌ها برای استفاده می‌باشد. لازم به توضیح است که زمان برداشت میوه سیب در منطقه مورد نظر اواخر شهریور و اوایل مهر ماه است. بنابراین با همکاری خبرگان محلی یک مجموعه داده با ۱۰۰ نمونه تهیه شده است.

فازی سازی مجموعه داده

در این بخش مقادیر عددی را به مقادیر فازی تبدیل می‌کنیم تا مقادیر ارزش خصیصه‌ها در مجموعه داده مشابه هم شود و برداشت ثانویه بهتر گردد. خصیصه‌های عددی تعداد روزهای تمام گل و میزان سفتی میوه بود که با مقادیر زیر فازی می‌گردد.

خصیصه میزان نرمی میوه

(میزان فرورفتگی به وسیله دستگاه فشارسنج به میلی‌متر)

- (۵–۷,۹۹) = نرم soft
- (۸–۹,۹۹) = متوسط medium
- (۱۰–۱۲) = سخت hard

تعداد روزهای تمام گل

- (۱۰۰–۱۱۵) = خیلی زود veryearly
- (۱۱۶–۱۲۹) = زود early
- (۱۳۰–۱۴۰) = بموقع timely

فرآیند انجام پروژه

در مرحله پیاده‌سازی و انجام پروژه از نرم‌افزار متن‌باز [1] وکا[2] به عنوان یک نرم‌افزار پیاده‌سازی یادگیری ماشین استفاده شده است. برای استفاده مجموعه داده در این نرم‌افزار باید آن در قالب فایل متنی[3] تهیه شود تا توسط نرم‌افزار وکا مورد استفاده قرار گیرد. سپس با اسفاده قابلیتهای محیط وکا ابزارهای مورد نیاز

1 Open Source
2 Weka
3 Text

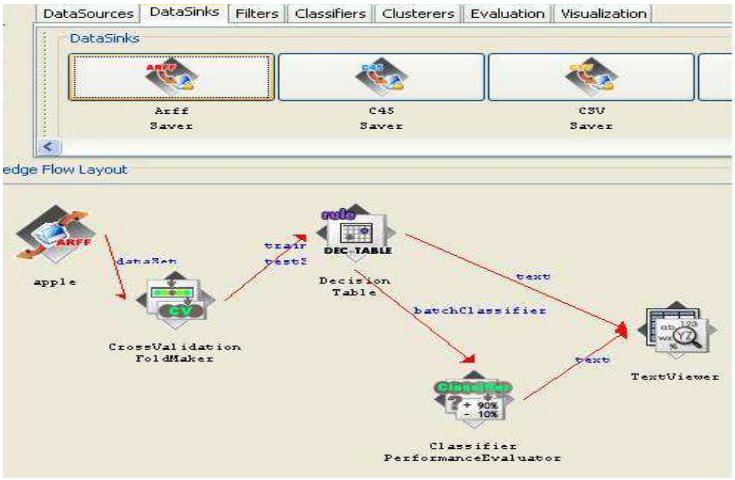
پروژه شامل الگوریتم یادگیری، خوشه‌بندی[1]، ابزار نمایش یا مصورسازی[2] و غیره را فراخوانی نموده ساختار اجرایی محیط پروژه را پیاده‌سازی می‌نماییم. سپس ارتباط بین اجزا را تعریف کرده و در ادامه مجموعه داده را فراخوانی نموده و پروژه را اجرا[3] می‌نمایم.

تبدیل مجموعه داده خام به فرمت نرم‌افزار وکا

یکی از فرمتها و قالبهای داده مورد استفاده وکا arff است که در این مرحله ما مجموعه داده خود را در یک فایل متنی به این فرمت تبدیل کرده و با پسوند arff. ذخیره می‌نماییم. Apple نام رابطه است که در ابتدا تعریف می‌شود و سپس خصیصه‌ها همراه با دامنه مقادیرشان ارائه می‌شوند و بعد از آن کلاس رابطه با مقادیر کلاس آورده می‌شوند. (لازم به توضیح است که در اینجا کلاس نیز مانند خصیصه‌ها و به عنوان آخرین خصیصه تعریف می‌شود). در نهایت هم مقادیر نمونه‌ها به ترتیب مشخص و بر اساس ترتیب تعریف خصیصه‌ها و بر اساس مجموعه داده جمع‌آوری شده تکمیل می‌شوند.

ساختار اجرایی محیط پروژه

در مرحله اجرا استفاده از قابلیتهای محیط وکا و ابزارهای آن مانند الگوریتم یادگیری، خوشه‌بندی، ابزار نمایش یا مصورسازی و غیره است. فرآیند کار بدین ترتیب است که ابتدا از گزینه Datasinks نوع فرمت مورد استفاده را تعیین می‌کنیم و با کلیک راست روی این آیکون و فراخوانی فایل apple.arff فایل مجموعه آموزشی را فراخوانی می‌نماییم. سپس الگوریتم یادگیری مورد نظر را از بخش classifiers انتخاب می‌نماییم و بعد از آن خروجی الگوریتم یادگیری روی مجموعه آموزشی را به یک classifier دیگر می‌دهیم و در نهایت نتیجه را به یک مصور ساز یا نمایشگر متنی[4] یا گرافیکی[5] برای نمایش دادن خروجی ارسال می‌نماییم. نمونه-ای از ساختار اجرایی محیط پروژه را در شکل ۲ می‌بینیم. پس از تعریف ارتباط بین اجزا در ادامه پروژه را به اجرا در می‌آوریم.



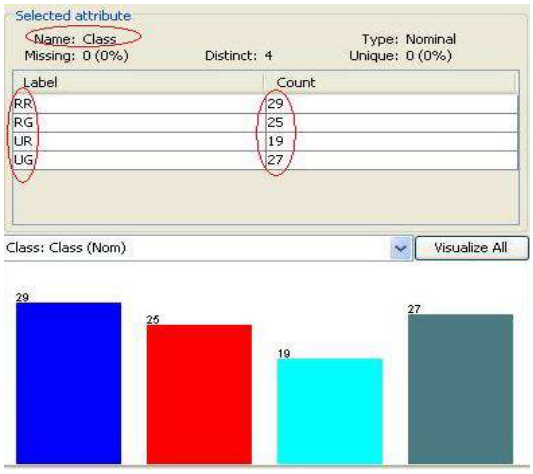
شکل۲: نمونه‌ای از ساختار اجرایی محیط پروژه

در شکل اجزای ساختار شامل ارتباط مجموعه آموزشی، الگوریتم یادگیری، خوشه‌بندی و نمایشگر خروجی مشاهده می‌شود.

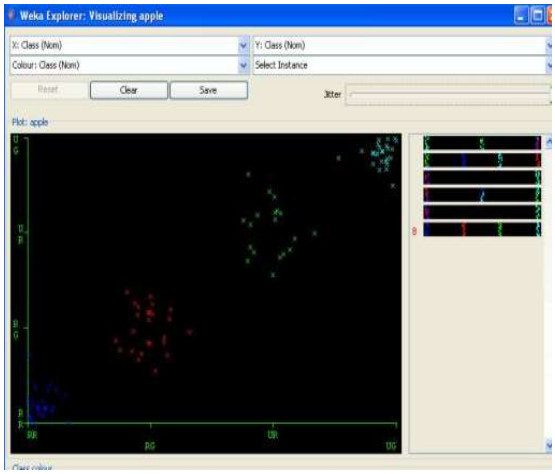
1 Clustering
2 Visualization
3 Run
4 TextViewer
5 GraphicViewer

مشاهده نتایج کلاسبندی

در این بخش نتایج پراکندگی نمونه‌ها نسبت به خصیصه‌ها در حالات مختلف نشان‌داده شده است. نتایج زیر نشان دهنـده فراوانـی کلاسها نسبت به هر خصیصه به صورت جداگانه و مقادیر ارزش خصیصه‌ها می‌باشد.
شکلهای ۳ و ۴ نتایج کلاسها نسبت به خود و تعداد فرآوانی آن نشان داده شده است.

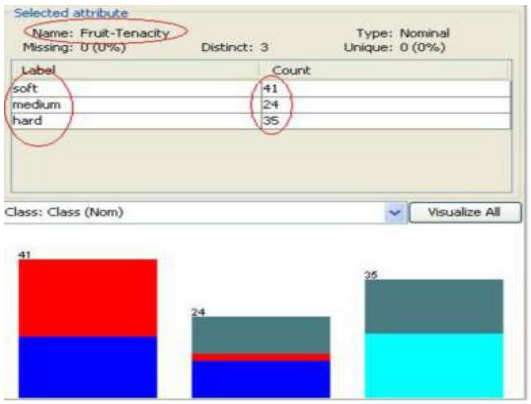


شکل۳: فراوانی خود کلاس

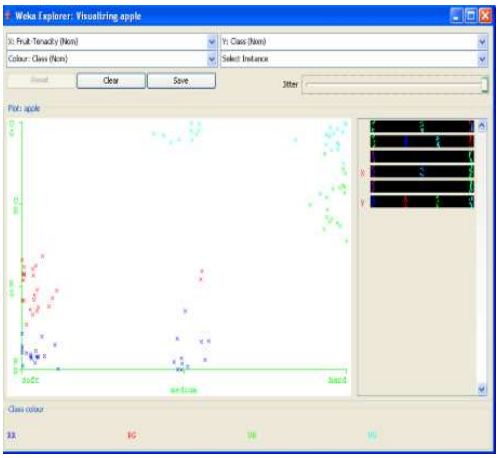


شکل۴: کلاسبندی بر اساس چگونگی پراکندگی خود کلاس

شکلهای ۵ و۶ نتایج کلاسها نسبت به میزان سفتی میوه و تعداد فرآوانی آن نشان داده شده است.

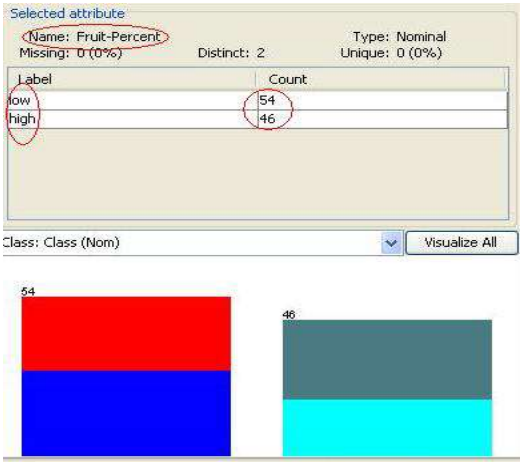


شکل ۵: فراوانی خصیصه میزان سفتی میوه

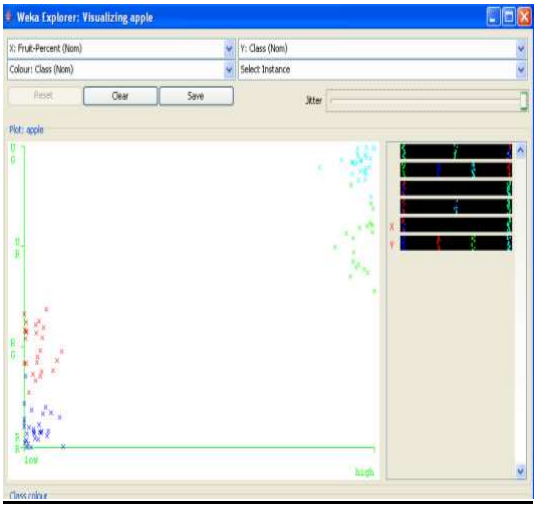


شکل ۶: کلاسبندی بر اساس چگونگی پراکندگی خصیصه میزان سفتی میوه

نوع پراکندگی نشان‌دهنده عدم تفکیک کلاسها بر اساس خصیصه‌های موجود است. بدین صورت رنگ قرمز و آبی (دو کـلاس) در بخش نرم با هم ادغام شده‌اند و در بخش متوسط نیز سه رنگ قرمز، آبی و سبز کم رنگ (سه کلاس) ادغام شـده انـد و دو رنـگ سبز کم رنگ و پر رنگ (دو کلاس) هم در بخش سخت قابل جداسازی نیستند.
شکلهای ۷ و ۸ نتایج کلاسها نسبت به میزان نشاسته و تعداد فرآوانی آن نشان داده شده است.

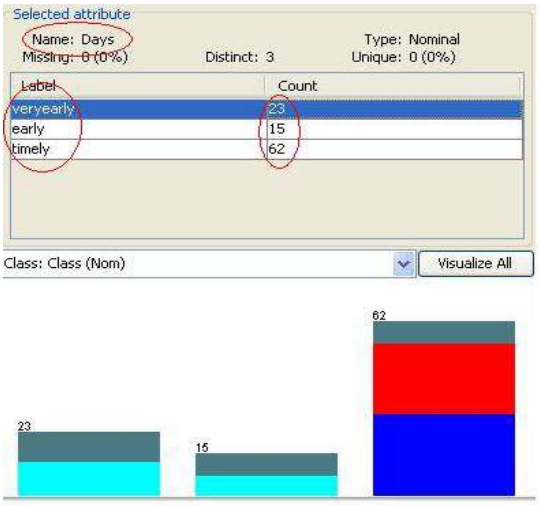


شکل ۷: فراوانی خصیصه میزان نشاسته

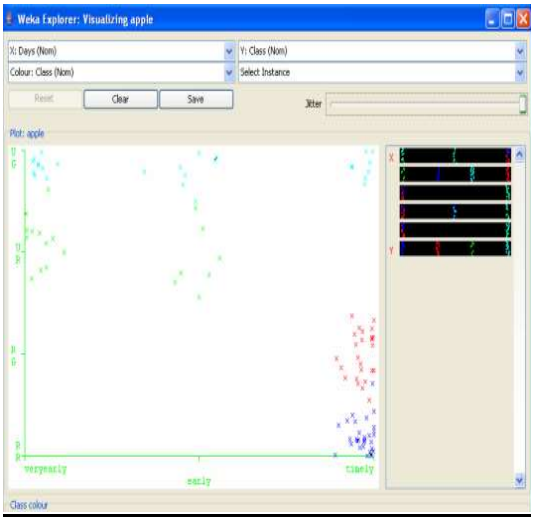


شکل ۸: کلاسبندی بر اساس چگونگی پراکندگی خصیصه میزان نشاسته

نوع پراکندگی نشان‌دهنده عدم تفکیک کلاسها مناسب بر اساس خصیصه موجود است. بدین صورت رنگ قرمز و آبی (دو کـلاس) در بخش کم تفکیک مناسبی ندارند و در بخش زیاد رنگ سبز کم رنگ و پر رنگ (دو کلاس) به راحتی قابـل جداسـازی نیـستند. خصیصه میزان نشاسته بهتر از خصیصه قبلی بوده ولی باز به عنوان خصیصه خوب مناسب نیست.
شکلهای ۹ و ۱۰ نتایج کلاسها نسبت به تعداد روزهای تمام گل و تعداد فرآوانی آن نشان داده شده است.



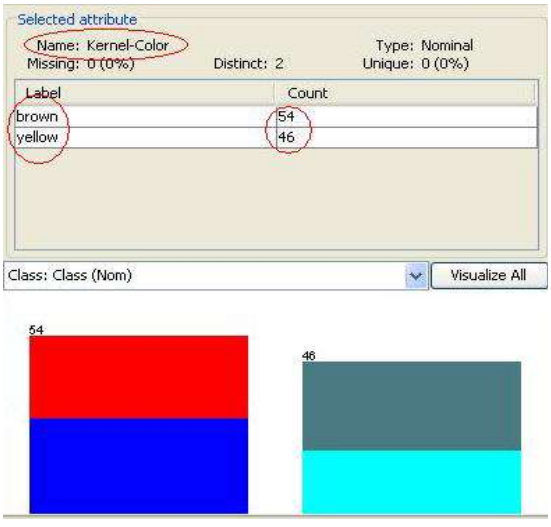
شکل ۹: فراوانی خصیصه تعداد روزهای تمام گل



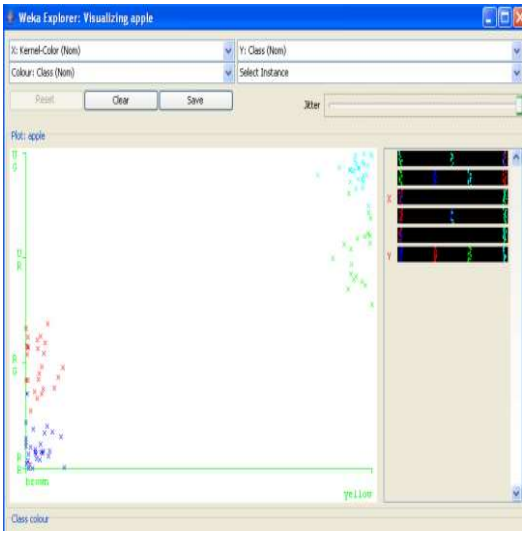
شکل ۱۰: کلاسبندی بر اساس چگونگی پراکندگی خصیصه تعداد روزهای تمام گل

در مورد روزهای تمام گل نیز نوع پراکندگی نشان‌دهنده عدم تفکیک کلاسها بر اساس خصیصه‌های موجود است. بدین صورت رنگ قرمز، و آبی و سبز پررنگ (سه کلاس) در مقدار بموقع با هم ادغام شده‌اند و در مقادیر بخش زود و خیلی زود دو رنگ سبز کم رنگ و پر رنگ (دو کلاس) در هر دو مقدار وجود دارد.

شکلهای ۱۱ و ۱۲ نتایج کلاسها نسبت به تعداد روزهای تمام گل و تعداد فرآوانی آن نشان داده شده است.

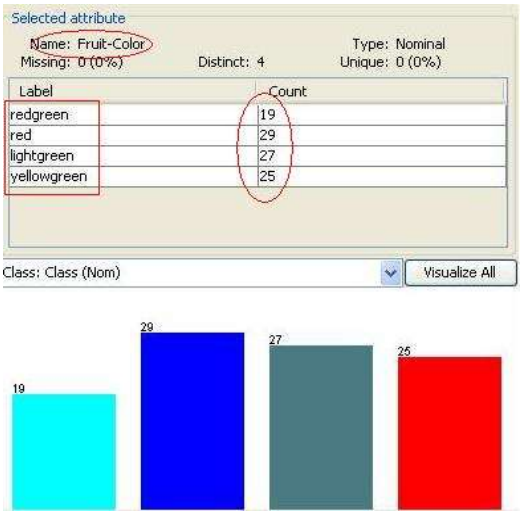


شکل ۱۱: فراوانی خصیصه رنگ هسته یا دون میوه

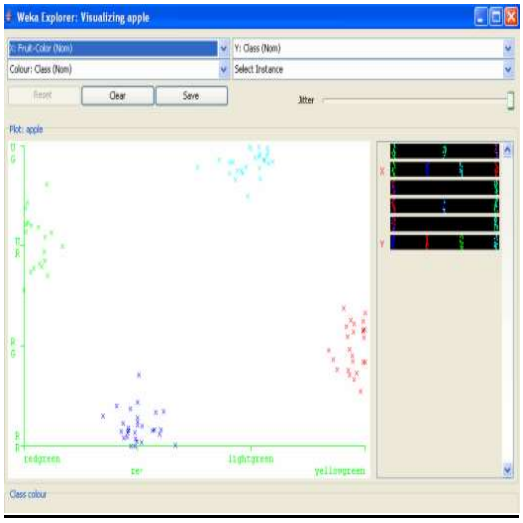


شکل ۱۲: کلاسبندی بر اساس چگونگی پراکندگی خصیصه رنگ هسته یا دون میوه

نوع پراکندگی نشان‌دهنده تفکیک نسبتا مناسب کلاسها بر اساس خصیصه‌های موجود است. بدین صورت رنگ قرمز و آبی در مقدار هسته قهوه‌ای با هم جدا شده‌اند و در بخش دو رنگ سبز کم رنگ و پر رنگ هم در مقدار هسته زرد تا حدودی جداسازی شده‌اند. ولی تفرق مانند خصیصه میزان نشاسته خوب نیست.
شکلهای ۱۳ و ۱۴ نتایج کلاسها نسبت به تعداد روزهای تمام گل و تعداد فرآوانی آن نشان داده شده است.



شکل۱۳: فراوانی خصیصه رنگ میوه



شکل ۱۴: کلاسبندی بر اساس چگونگی پراکندگی خصیصه رنگ میوه

نوع پراکندگی نشان‌دهنده تفکیک بسیار مناسب کلاسها بر اساس مقادیر خصیصه‌های موجود است. بدین صورت رنگ قرمز که نشان دهنده کلاس طلایی رسیده است، کاملاً در مقدار یک خصیصه یعنی زرد مایل به سبز و رنگ آبی که نشان دهنده کلاس قرمز رسیده است، کاملاً در مقدار یک خصیصه یعنی قرمز قرار دارد. همچنین رنگ سبز که نشان دهنده کلاس طلایی نارس است، کاملاً در مقدار یک خصیصه یعنی سبز روشن و رنگ سبز پررنگ که نشان دهنده کلاس قرمز نارس رسیده است، کاملاً در مقدار یک خصیصه یعنی سبز با رگه‌های قرمز قرار دارد.

نتیجه

نتایج پراکندگی نمونه‌ها نشان دهنده این است که در مجموعه داده جمع آوری شده خصیصه رنگ میوه کلاسبندی مناسبی انجام می‌دهد و نتایج کلاسبندی با این خصیصه، مرزهای مشخص داشته و جداسازی و تفرق مناسبی بین نمونه‌ها و کلاسهای آنها صورتگرفته است. شکل ۱۴ پراکندگی نمونه‌ها و کلاس آنها را نسبت به خصیصه رنگ و ارزش و مقدار آن نشان داده است که به راحتی نمونه‌ها از هم جدا شده‌اند و در شکلهای ۱۲ و ۸ جداسازی و مرز بین نمونه‌ها مناسب نیست. در شکلهای ۶ و ۱۰ که کلاسبندی بر اساس خصیصه میزان سفتی میوه و روزهای تمام گل است نیز امکان جداسازی نمونه‌ها به خوبی وجود ندارد و دو خصیصه روزهای تمام گل و میزان سفتی میوه شرایط مشابه یعنی مشخص نبودن مرزبندی بین کلاسها را دارا بودند. بنابراین در این مجموعه داده و نمونه‌های جمع‌آوری شده، خصیصه رنگ میوه نسبت به سایر خصیصه‌ها نتیجه بهتری داشته است.

سپاسگزاری

با تشکر از همکاری صمیمانه سازمان جهاد کشاورزی استان آذربایجان غربی خصوصاً جناب آقای مهندس پرویز حیدریزاد رییس محترم دفتر آمار و فناوری اطلاعات سازمان جهاد کشاورزی استان آذربایجان غربی و کارشناسان دفتر مدیریت باغبانی که در جمع آوری اطلاعات و بازدید میدانی از باغات سیب استان ما را یاری نموده‌اند.