

سوال ۱-

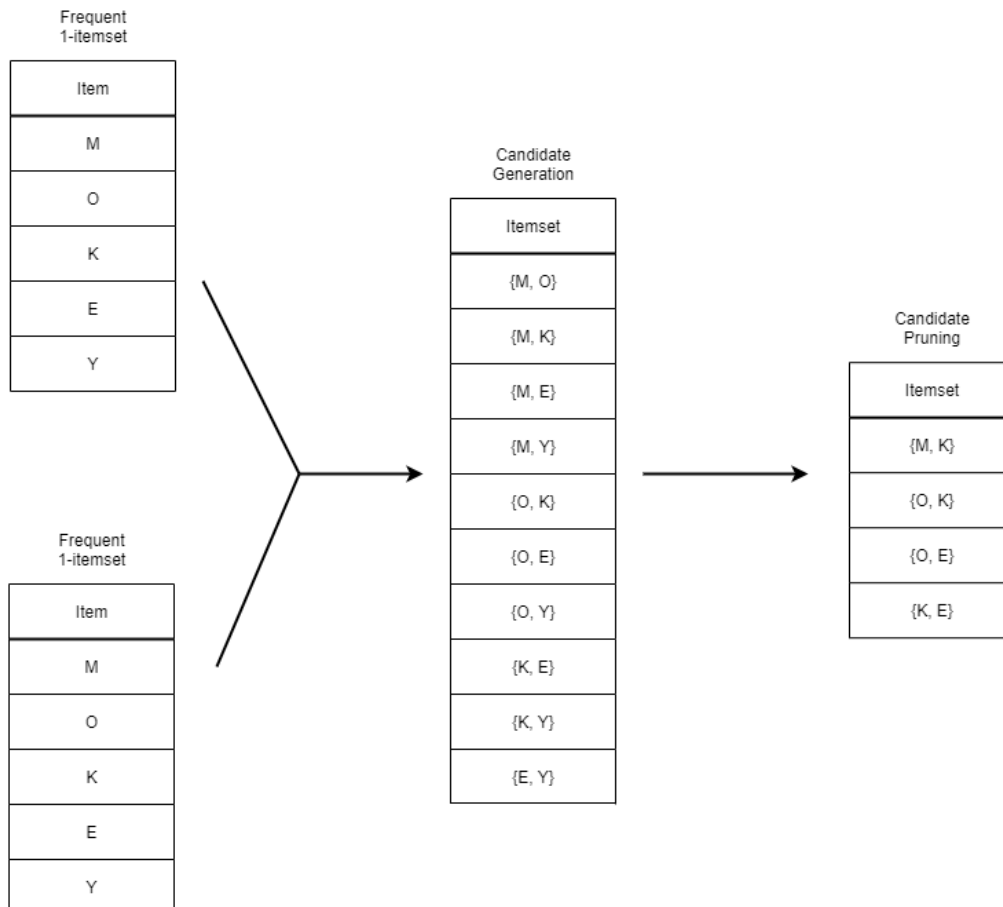
در دسته‌بندی تنبل تنها داده‌های آموزش را ذخیره می‌کنیم و صبر می‌کنیم تا یک داده تست ظاهر شود. در این هنگام دسته‌بندی را بر اساس مرتبط‌ترین داده میان داده‌های آموزشی ذخیره شده انجام می‌دهیم. در مقایسه با دسته‌بندی مشتاق، در دسته‌بندی تنبل زمان کمتری صرف آموزش می‌شود اما پیش‌بینی زمان بیشتری می‌گیرد. مانند: KNN

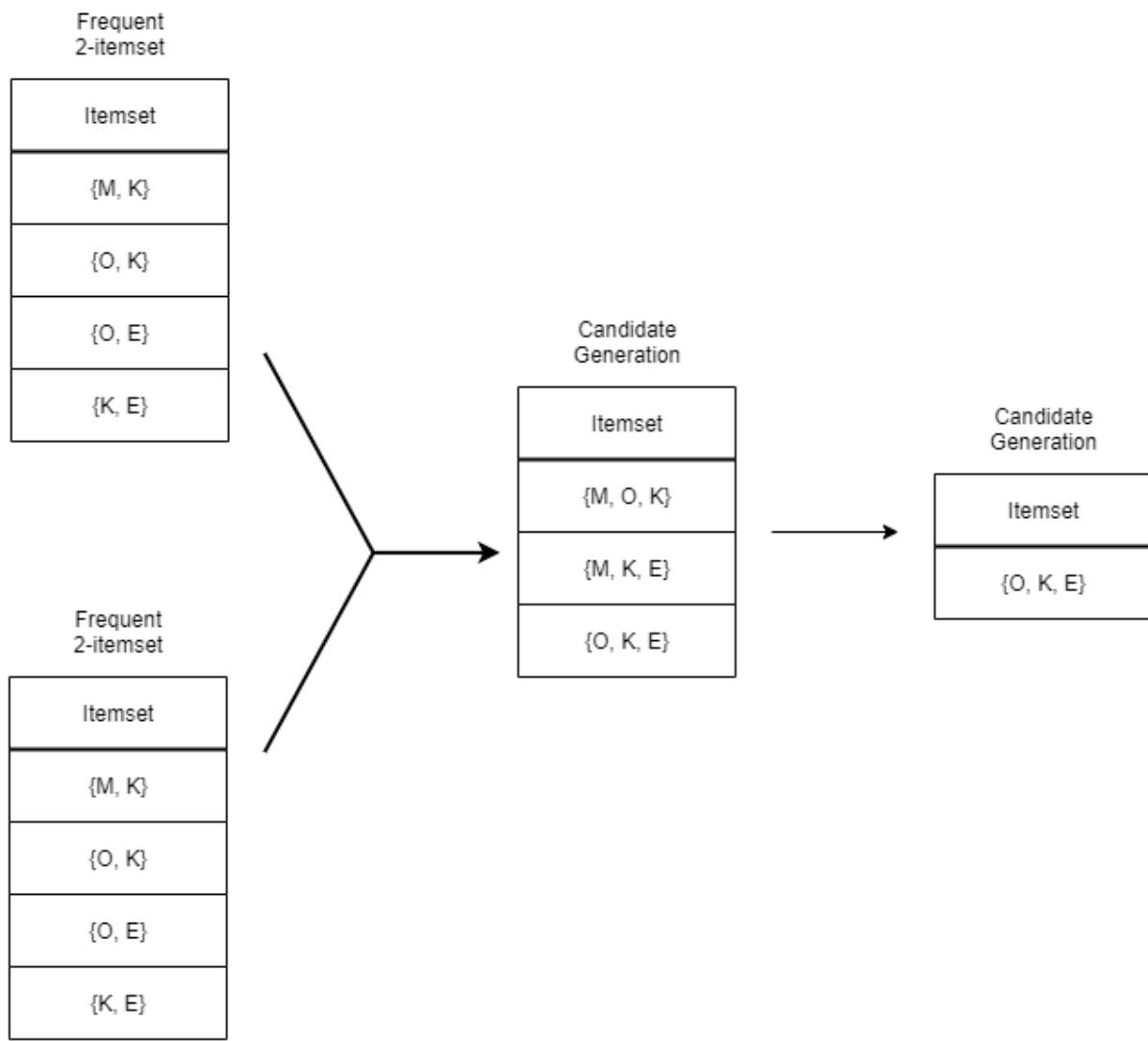
در دسته‌بندی مشتاق قبل از اینکه داده‌ای برای دسته‌بندی دریافت کنیم، یک مدل دسته‌بندی بر اساس داده‌های آموزش که در اختیارمان قرار دارد می‌سازیم. این مدل به یک فرض واحد برای کل فضای نمونه متعهد می‌باشد. ساخت مدل باعث می‌شود که دسته‌بندی مشتاق زمان بیشتری برای آموزش و زمان کمتری برای پیش‌بینی مصرف کند. مانند: درخت تصمیم‌گیری، Naïve Bayes و شبکه عصبی

سوال ۲-

(الف)

Apriori Algorithm:





Frequent 1-itemset

Item
M
O
K
E
Y

Frequent 2-itemset

Itemset
{M, K}
{O, K}
{O, E}
{K, E}

Frequent 3-itemset

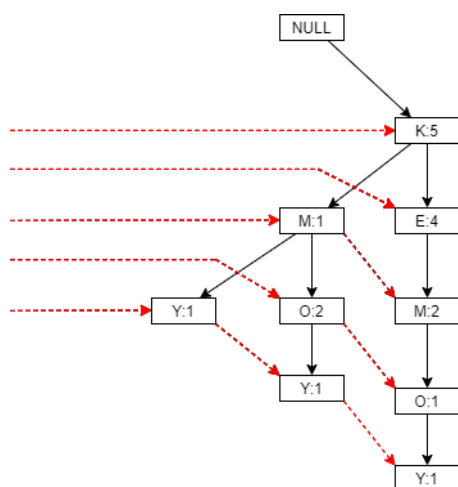
Itemset
{O, K, E}

FP-growth Algorithm:

ابتدا آیتم‌های کم تکرار را حذف می‌کنیم و آن‌ها را برحسب پرتکرار بودن مرتب می‌کنیم:

TID	Items_bought	Sorted_items
T1	{M, O, N, K, E, Y}	{K, E, M, O, Y}
T2	{D, O, N, K, E, Y}	{K, E, O, Y}
T3	{M, A, K, E}	{K, E, M}
T4	{M, U, C, K, Y}	{K, M, Y}
T5	{C, O, O, K, I, E}	{K, E, O}

حال آیتم‌های مرتب شده را به درخت اضافه می‌کنیم:



حال مسیرهای منتهی به آیتم‌ها را بدست می‌آوریم:

Items	Conditional Pattern Base
Y	{{K, E, M, O : 1}, {K, E, O : 1}, {K, M : 1}}
O	{{K, E, M : 1}, {K, E : 2}}
M	{{K, E : 2}, {K : 1}}
E	{K : 4}
K	-

حال با جدا کردن قسمت مشترک مسیرها و اضافه کردن آیتم انتهایی مسیر frequent itemset بدست می‌آیند:

Items	Frequent Pattern Generated
Y	{<K, Y : 3>}
O	{<K, O : 3>, <E, O : 3>, <E, K, O : 3>}
M	{<K, M : 3>}
E	{<E, K : 3>}
K	-

(ب)

هر دو الگوریتم **Apriori** و **FP Growth** برای استخراج پترن‌های پرتکرار در دیتابیس‌ها استفاده می‌شوند و هر دوی آن‌ها از تکنیک‌هایی برای کشف این پترن‌ها استفاده می‌کنند. الگوریتم **Apriori** با دیتابیس‌های بزرگ خوب کار می‌کند در صورتی که الگوریتم **FP Growth** کارایی زیادی در دیتابیس‌های بزرگ ندارد. در ادامه این دو الگوریتم را در زمینه‌های مختلف با هم مقایسه می‌کنیم:

- **تکنیک:** الگوریتم **Apriori** از خاصیت **Apriori** و ادغام کردن **itemlist**ها استفاده می‌کند. الگوریتم **FP Growth** پترن‌های شرطی می‌سازد که حداقل **support** را ارضا کنند.
- **نوع جستجو:** **Apriori** از **BFS** و **FP Growth** از **Divide & Conquer** استفاده می‌کند.
- **حافظه:** الگوریتم **Apriori** مقدار زیادی از فضای حافظه را مصرف می‌کند زیرا با تعداد زیادی **itemset** کاندید سر و کار دارد. الگوریتم **FP Growth** به دلیل ساختار فشرده‌اش حافظه کمتری لازم دارد.
- **خواندن دیتابیس:** الگوریتم **Apriori** برای ساختن مجموعه کاندید چندین بار دیتابیس را می‌خواند در صورتی که الگوریتم **FP Growth** دیتابیس را تنها دوبار می‌خواند.
- **زمان:** در الگوریتم **Apriori** زمان بیشتری برای هر بار ساختن کاندیدها تلف می‌شود. زمان اجرای **FP Growth** در مقایسه با **Apriori** کمتر است.

سوال ۳-

یادگیری نیمه‌نظارتی بین یادگیری نظارتی و غیرنظارتی قرار می‌گیرد. در نوع مسئله، بخش کوچکی از نمونه‌ها برچسب دارند و ما باید تلاش کنیم که از بقیه نمونه‌های بدون برچسب به گونه‌ای استفاده کنیم تا نسبت به زمانی یادگیری نظارتی و غیرنظارتی را به صورت جداگانه انجام می‌دهیم، کارایی بیشتری بدست آوریم.

یادگیری فعالانه نوع خاصی از یادگیری نیمه‌نظارتی می‌باشد. در اینجا ما از نمونه‌های برچسب‌دار استفاده می‌کنیم و در میان نمونه‌های بدون برچسب باید یک برچسب‌گذاری پیدا کنیم که با تعداد کمی برچسب بیشترین کارایی را بدست آوریم. این یک الگوریتم تکرارشونده می‌باشد که در آن یادگیرنده باید در هر مرحله از کاربر بخواهد که تعدادی از نمونه‌ها را برچسب‌گذاری کند و این فرآیند تا زمان همگرایی تکرار می‌شود. به این روش فقط به تعداد کمی برچسب‌گذاری نیاز داریم تا به کارایی بالا برسیم.

یادگیری انتقالی به معنای استفاده از راه حل یک مسئله و سازگار کردن آن با مسئله هدف می‌باشد. از این نوع یادگیری در **Deep Learning** استفاده می‌شود. به طور معمول از دیتاست **imagenet** برای یادگیری اولیه شبکه استفاده می‌شود و پس از آن با توجه به مسئله اصلی تغییرات در آن اعمال می‌شود.

سوال ۴-

دو دسته اصلی روش‌های یادگیری گروهی **bagging** و **boosting** می‌باشند.

Bagging یکی از اولین و ساده‌ترین روش‌های یادگیری گروهی می‌باشد و به طرز شگفت‌انگیزی کارایی بالایی دارد. تنوع در دسته‌بندی‌های روش **bagging** با استفاده از روش **bootstrap** و ساخت نسخه‌های مختلف از داده‌های یادگیری بدست می‌آید. به این صورت که زیرمجموعه‌های تصادفی مختلفی از داده‌های یادگیری انتخاب می‌کنیم و با هر کدام از آن‌ها یک **classifier** متفاوت را آموزش می‌دهیم. آنگاه

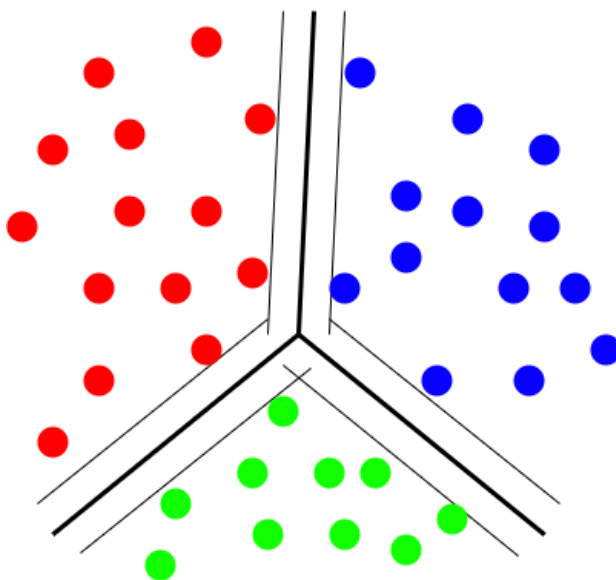
classifierها را با یکدیگر ترکیب می‌کنیم. این کار با یک رای‌گیری ساده میان تصمیم‌گیری‌های آن‌ها انجام می‌شود. در واقع کلاسی که توسط بیشتر classifierها انتخاب شود انتخاب نهایی گروه می‌باشد.

در boosting نیز مشابه با bagging یک گروه از classifierها را با چند بار نمونه‌گیری از داده‌ها ایجاد می‌کنیم که نهایتاً با رای‌گیری با هم ترکیب می‌شوند. اما در boosting نمونه‌برداری به گونه‌ای استراتژیک انجام می‌شود تا اطلاعات بیشتری را در اختیار classifier بعدی قرار دهد. یک مثال می‌تواند به این گونه باشد که هر تکرار از boosting سه classifier ضعیف بسازد: اولین classifier (C_1) با یک زیرمجموعه تصادفی از داده‌های آموزشی موجود آموزش داده می‌شود. داده‌های آموزشی برای classifier دوم (C_2) از زیرمجموعه‌ای انتخاب می‌شود که شامل بیشترین اطلاعات برای C_1 بوده است. به طور دقیق‌تر، C_2 با استفاده از داده‌هایی آموزش داده می‌شود که تنها نیمی از آن‌ها توسط C_1 به درستی دسته‌بندی شده‌اند و باقی آن‌ها به اشتباه دسته‌بندی شده است. سومین classifier (C_3) بر روی داده‌هایی آموزش داده می‌شود که C_1 و C_2 بر سر آن‌ها اختلاف نظر دارند. این سه classifier توسط رای‌گیری با یکدیگر ترکیب می‌شوند.

سوال ۵-

(الف)

یک سناریو از SVM سه کلاسه را در شکل زیر مشاهده می‌کنید:



در واقع به جای دو margin در اینجا سه margin داریم که باید آن‌ها را بیشینه کنیم.

(ب)

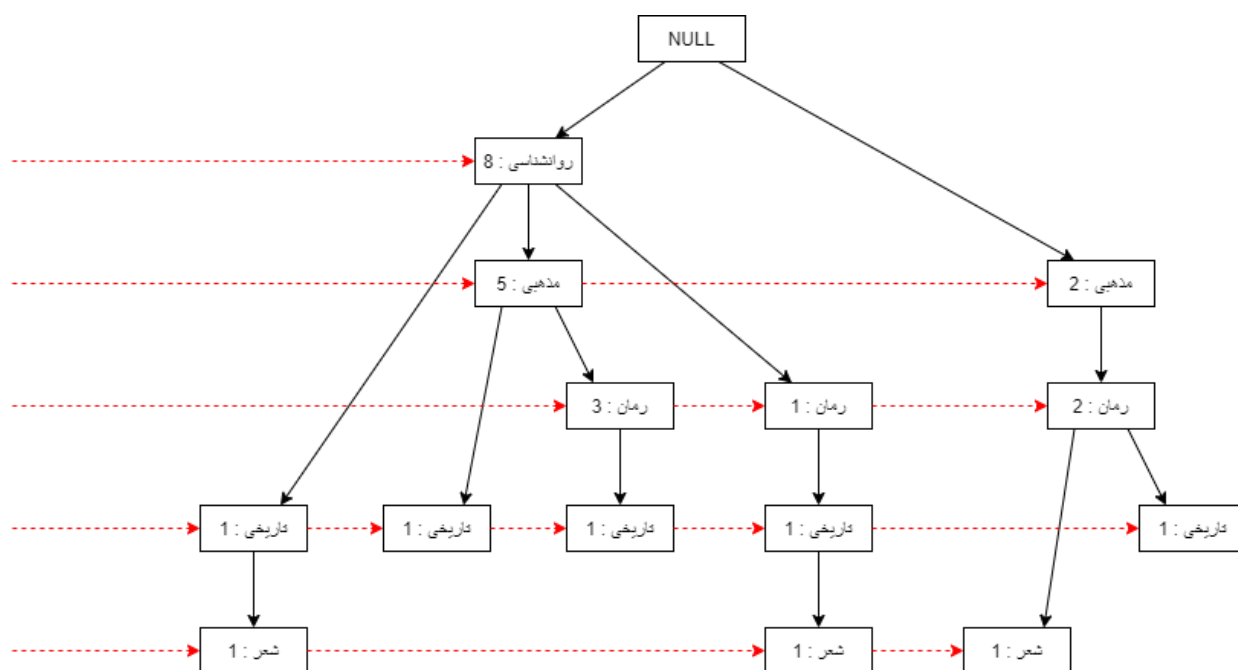
منظور از حاشیه سخت این است که می‌خواهیم همه داده‌ها به درستی دسته‌بندی شوند و هیچ داده‌ای در margin قرار نگیرد. در این حالت به اصطلاح محدودیت‌ها سخت‌گیرانه بوده و برای outlierها استثناء در نظر گرفته نمی‌شود. اگر داده‌های ما linearly separable نباشند، امکان استفاده از SVM با حاشیه سخت نمی‌باشد.

ج)

دلیل کم برآزش احتمالا این است که داده‌های ما به صورت خطی جداپذیر نیستند. برای حل این مشکل باید پارامترهای مدل را به یک فضای برداری جدید ببریم تا بتوانیم در فضای جدید با یک مرز تصمیم‌گیری خطی داده‌ها را از همدیگر جدا کنیم. برای این منظور یک تابع kernel تعریف می‌کنیم به صورتی که تبدیلی وجود داشته باشد که مقدار تابع kernel محاسبه شده برای دو بردار برابر با حاصل ضرب داخلی آن دو بردار در فضای جدید باشد.

سوال ۶-

آیتم‌ها در لیست‌های داده شده بر حسب تعداد تکرار مرتب شده‌اند و نیازی به مرتب سازی آن‌ها نیست. پس آن‌ها را مستقیما به درخت اضافه می‌کنیم:



حال مسیرهای منتهی به آیتم‌ها را بدست می‌آوریم:

Items	Conditional Pattern Base
شعر	{ {مذهبی، رمان: ۱}، {روانشناسی، رمان، تاریخی: ۱}، {روانشناسی، تاریخی: ۱} }
تاریخی	{ {مذهبی، رمان: ۱}، {روانشناسی، رمان: ۱}، {روانشناسی، مذهبی، رمان: ۱}، {روانشناسی، مذهبی: ۱}، {روانشناسی: ۱} }
رمان	{ {مذهبی: ۲}، {روانشناسی: ۱}، {روانشناسی، مذهبی: ۳} }
مذهبی	{ {روانشناسی: ۵} }
روانشناسی	-

حال با ترکیب کردن مسیرها، اضافه کردن انتهای مسیر و نگه داشتن آن‌هایی که حداقل پشتیبانی را دارند، frequent itemset ها را بدست می‌آوریم:

Items	Frequent Pattern Generated
شعر	{(رمان، شعر:۲)، (روانشناسی، شعر:۲)، (تاریخی، شعر:۲)، (روانشناسی، تاریخی، شعر:۲)}
تاریخی	{(مذهبی، تاریخی:۳)، (رمان، تاریخی:۳)، (روانشناسی، تاریخی:۴)، (مذهبی، رمان، تاریخی:۲)، (روانشناسی، رمان، تاریخی:۲)}
رمان	{(مذهبی، رمان:۵)، (روانشناسی، رمان:۴)، (روانشناسی، مذهبی، رمان:۳)}
مذهبی	{(روانشناسی، مذهبی:۵)}
روانشناسی	-

(ب)

همه قواعد ممکن را از frequent itemset های بدست آمده می‌سازیم و confidence آن‌ها را محاسبه می‌کنیم:

$$\text{شعر} \rightarrow \text{رمان} = \frac{2}{6} = 33\%$$

$$\text{شعر} \rightarrow \text{رمان} = \frac{2}{3} = 67\%$$

$$\text{شعر} \rightarrow \text{روانشناسی} = \frac{2}{8} = 25\%$$

$$\text{شعر} \rightarrow \text{روانشناسی} = \frac{2}{3} = 67\%$$

$$\text{شعر} \rightarrow \text{تاریخی} = \frac{2}{5} = 40\%$$

$$\text{شعر} \rightarrow \text{تاریخی} = \frac{2}{3} = 67\%$$

$$\text{شعر} \& \text{تاریخی} \rightarrow \text{روانشناسی} = \frac{2}{8} = 25\%$$

$$\text{شعر} \& \text{روانشناسی} \rightarrow \text{تاریخی} = \frac{2}{5} = 40\%$$

$$\text{شعر} \& \text{تاریخی} \rightarrow \text{روانشناسی} = \frac{2}{3} = 67\%$$

$$\text{شعر} \rightarrow \text{تاریخی} \& \text{روانشناسی} = \frac{2}{4} = 50\%$$

$$\text{تاریخی} \rightarrow \text{شعر} \& \text{روانشناسی} = \frac{2}{2} = 100\%$$

$$\text{روانشناسی} \rightarrow \text{شعر} \& \text{تاریخی} = \frac{2}{2} = 100\%$$

$$\text{تاریخی} \rightarrow \text{مذهبی} = \frac{3}{7} = 43\%$$

$$\text{مذهبی} \rightarrow \text{تاریخی} = \frac{3}{5} = 60\%$$

$$\text{تاریخی} \rightarrow \text{رمان} = \frac{3}{6} = 50\%$$

$$\text{رمان} \rightarrow \text{تاریخی} = \frac{3}{5} = 60\%$$

$$\text{روانشناسی} \rightarrow \text{تاریخی} = \frac{4}{8} = 50\%$$

$$\text{تاریخی} \rightarrow \text{روانشناسی} = \frac{4}{5} = 80\%$$

$$\text{مذهبی} \rightarrow \text{رمان} \& \text{تاریخی} = \frac{2}{7} = 29\%$$

$$\text{رمان} \rightarrow \text{مذهبی} \& \text{تاریخی} = \frac{2}{6} = 33\%$$

$$\text{تاریخی} \rightarrow \text{رمان} \& \text{مذهبی} = \frac{2}{5} = 40\%$$

$$\text{مذهبی} \rightarrow \text{رمان} \& \text{تاریخی} = \frac{2}{5} = 40\%$$

$$\text{رمان} \rightarrow \text{تاریخی} \& \text{مذهبی} = \frac{2}{3} = 67\%$$

$$\text{مذهبی} \rightarrow \text{تاریخی} \& \text{رمان} = \frac{2}{3} = 67\%$$

$$\text{روانشناسی} \rightarrow \text{رمان} \& \text{تاریخی} = \frac{2}{8} = 25\%$$

$$\text{رمان} \rightarrow \text{روانشناسی} \& \text{تاریخی} = \frac{2}{6} = 33\%$$

$$\text{تاریخی} \rightarrow \text{رمان} \& \text{روانشناسی} = \frac{2}{5} = 40\%$$

$$\text{روانشناسی} \rightarrow \text{رمان} \& \text{تاریخی} = \frac{2}{4} = 50\%$$

$$\text{رمان} \rightarrow \text{تاریخی} \& \text{روانشناسی} = \frac{2}{4} = 50\%$$

$$\text{روانشناسی} \rightarrow \text{تاریخی} \& \text{رمان} = \frac{2}{3} = 67\%$$

$$\text{مذهبی} \rightarrow \text{رمان} = \frac{5}{7} = 71\%$$

$$\text{رمان} \rightarrow \text{مذهبی} = \frac{5}{6} = 83\%$$

$$\text{روانشناسی} \rightarrow \text{رمان} = \frac{4}{8} = 50\%$$

$$\text{رمان} \rightarrow \text{روانشناسی} = \frac{4}{6} = 67\%$$

$$\text{روانشناسی} \rightarrow \text{رمان} \& \text{مذهبی} = \frac{3}{8} = 38\%$$

$$\text{مذهبی} \rightarrow \text{رمان} \& \text{روانشناسی} = \frac{3}{7} = 43\%$$

$$\text{رمان} \rightarrow \text{مذهبی} \& \text{روانشناسی} = \frac{3}{6} = 50\%$$

$$\text{روانشناسی} \rightarrow \text{رمان} = \frac{3}{5} = 60\%$$

$$\text{مذهبی} \rightarrow \text{رمان} \& \text{روانشناسی} = \frac{3}{4} = 75\%$$

$$\text{روانشناسی} \rightarrow \text{رمان} \& \text{مذهبی} = \frac{3}{5} = 60\%$$

$$\text{روانشناسی} \rightarrow \text{مذهبی} = \frac{5}{8} = 63\%$$

$$71\% = \frac{5}{7} - \text{روانشناسی} \rightarrow \text{مذهبی}$$

حال قواعد با حداقل confidence را جدا می‌کنیم:

رمان \rightarrow شعر

روانشناسی \rightarrow شعر

تاریخی \rightarrow شعر

تاریخی & روانشناسی \rightarrow شعر

شعر \rightarrow تاریخی & روانشناسی

تاریخی \rightarrow شعر & روانشناسی

روانشناسی \rightarrow شعر & تاریخی

مذهبی \rightarrow تاریخی

تاریخی \rightarrow رمان

رمان \rightarrow تاریخی

تاریخی \rightarrow روانشناسی

روانشناسی \rightarrow تاریخی

رمان \rightarrow تاریخی & مذهبی

مذهبی \rightarrow تاریخی & رمان

تاریخی \rightarrow رمان & روانشناسی

رمان \rightarrow تاریخی & روانشناسی

روانشناسی \rightarrow تاریخی & رمان

رمان \rightarrow مذهبی

مذهبی \rightarrow رمان

رمان \rightarrow روانشناسی

روانشناسی \rightarrow رمان

مذهبی & روانشناسی \rightarrow رمان

رمان \rightarrow مذهبی & روانشناسی

مذهبی \rightarrow رمان & روانشناسی

روانشناسی \rightarrow رمان & مذهبی

مذهبی \rightarrow روانشناسی

روانشناسی \rightarrow مذهبی

پیاده‌سازی:

سوال ۸-

```
criterion='gini', max_depth=3
Accuracy of classifier 1: 0.8707865168539326

criterion='gini', max_depth=10
Accuracy of classifier 2: 0.8764044943820225

criterion='entropy', max_depth=3
Accuracy of classifier 3: 0.8370786516853933

criterion='entropy', max_depth=10
Accuracy of classifier 4: 0.8764044943820225
```

-۱

بالاترین دقت بدست آمده از مدل جنگل تصادفی برابر با ۸۷ درصد می‌باشد. در صورتی که بالاترین دقت بدست آمده از مدل درخت تصمیم ۸۰ درصد بود.

-۲

سرعت یادگیری جنگل تصادفی حدود ۷۰ برابر کندتر از درخت تصمیم‌گیری می‌باشد زیرا باید چندین درخت را آموزش دهد. همچنین سرعت تست جنگل تصادفی حدود ۱۸ برابر کندتر از درخت تصمیم‌گیری است.

سوال ۹-

-۲

```
Accuracy of linear SVM: 0.797752808988764
```

-۳

دقت بدست آمده هم از درخت تصمیم‌گیری و هم از جنگل تصادفی کمتر می‌باشد زیرا داده‌های ما linearly separable نیستند. بنابراین مدلی که آموزش داده شده قابلیت جداسازی داده‌های با کلاس‌های متفاوت را ندارد.

-۴

```
Accuracy of non-linear SVM: 0.6573033707865169
```

-۵

در حالت غیر خطی دقت باز هم کاهش یافت. احتمالاً به این علت است که مدل دچار بیش‌برازش یا overfit شده است.