



دانشگاه پیام نور  
دانشکده فنی و مهندسی

پایان نامه

برای دریافت درجه کارشناسی ارشد  
در رشته مدیریت و فن آوری اطلاعات و ارتباطات  
گروه مهندسی کامپیوتر و فن آوری اطلاعات و ارتباطات

**عنوان :**

**بکارگیری الگوریتم‌های داده کاوی جهت دسته بندی مشتریان اعتباری بانک در  
وضعیت برچسب گذاری با استفاده از تکنیک های AHP و TOPSIS براساس  
شاخص های تعیین کننده ی کلاس مشتری**

**استاد راهنما :**

دکتر مهرداد جلالی

**استاد مشاور :**

دکتر داوود کریم زادگان مقدم

**نگارش :**

افسانه حاجی زاده

شهریور ۹۱

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه پیام نور  
دانشکده فنی و مهندسی

پایان نامه

برای دریافت درجه کارشناسی ارشد  
در رشته مدیریت و فن آوری اطلاعات و ارتباطات  
گروه مهندسی کامپیوتر و فن آوری اطلاعات و ارتباطات

**عنوان :**

**بکارگیری الگوریتم‌های داده کاوی جهت دسته بندی مشتریان اعتباری بانک در  
وضعیت برچسب گذاری با استفاده از تکنیک های AHP و TOPSIS براساس  
شاخص های تعیین کننده ی کلاس مشتری**

**استاد راهنما :**

دکتر مهرداد جلالی

**استاد مشاور :**

دکتر داود کریم زادگان مقدم

**نگارش :**

افسانه حاجی زاده

شهریور ۹۱

تقدیم به:

پدر، مادر و همسر عزیزم

## تشکر و قدردانی

بنام خداوندی که داشتن او جبران همه نداشته های من است می ستایمش، چون لایق ستایش است. در اینجا بر خود لازم دانسته که نهایت سپاس و قدردانی را نسبت به تمام عزیزانی که در انجام این پژوهش از راهنمائی ها و مساعدت های با ارزش و حمایت های معنوی آن ها، بهره مند بوده ام را ابراز نموده و توفیق روز افزونشان را از درگاه احدیت آرزو نمایم.

از زحمات جناب آقای دکتر مهرداد جلالی که به عنوان استاد راهنما از رهنمودها و کمک های بی دریغ ایشان بهره مند بوده ام تشکر و قدردانی می کنم. همچنین جا دارد از زحمات استاد محترم مشاور جناب آقای دکتر داوود کریم زادگان مقدم، سپاسگزاری نمایم.

## چکیده:

پایه ریزی و ایجاد یک سیستم برای کنترل ریسک مشتریان، جزئی ضروری از مدیریت علمی یک بانک به شمار می‌رود. نظر به اهمیت ریسک اعتباری در سیستم بانکی، مدل‌های امتیازدهی اعتباری به عنوان ابزارهایی جهت کنترل این ریسک، توسعه یافته‌اند.

در این پژوهش از ابزارهای داده کاوی و روش‌های تصمیم‌گیری چند معیاره AHP و TOPSIS برای رتبه بندی اعتباری مشتریان یک بانک استفاده شده است. هدف آنست که در نهایت مدل تصمیم‌گیری ارائه شده به شناسایی مشتریان قبل از اعطای وام کمک نماید. روش کار بدین شکل می‌باشد که طبق یک فرآیند استاندارد داده کاوی (CRISP-DM)، داده‌های مشتریان سابق بانک جمع‌آوری، متغیرهای تأثیرگذار در رفتار اعتباری مشتریان شناسایی و فرآیند پالایش بر روی داده‌ها صورت گرفته است. سپس بر اساس چند شاخص (از قبیل روزهای تأخیر، تعداد تأخیر و مبلغ تأخیر در بازپرداخت اقساط وام)، طبق نظر کارشناسان بانک مورد مطالعه و با استفاده از روش‌های تصمیم‌گیری چند معیاره AHP و TOPSIS، طبقات اعتباری مشتریان (خوب، متوسط و بد) تعیین گردند. به عبارتی برای هر مشتری یک برچسب اختصاص می‌یابد. پس از پیش پردازش داده‌ها و تغییرات در رکوردها (مشتریان) و فیلدها (خصیصه‌های مشتریان)، با استفاده از روش CART، داده‌های نهایی مدل سازی گردیدند و الگوهای پنهان در داده‌ها به شکل درخت قوانین استخراج شدند. همچنین داده‌های نهایی با چندین الگوریتم داده کاوی نظیر درخت‌های تصمیم سستی، شبکه‌های عصبی، رگرسیون لجستیک، شبکه‌های بیزین، ماشین بردار پشتیبان و ... نیز مدل سازی شدند. نتیجه حاصله نشان داد که الگوریتم CART نتایج بهتری را به لحاظ دقت تفکیک مشتریان نسبت به سایرین دارد.

## واژه‌های کلیدی:

رتبه بندی اعتباری، طبقه بندی، داده کاوی، الگوریتم CART، تصمیم‌گیری چند معیاره، AHP،

.TOPSIS

عنوان	صفحه
<b>فصل ۱ کلیات تحقیق</b>	
۱-۱. مقدمه.....	۲
۲-۱. بیان مسئله تحقیق.....	۳
۳-۱. سابقه و ضرورت تحقیق.....	۶
۱-۳-۱. جنبه جدید و نوآوری در فرآیند تحقیق.....	۶
۴-۱. اهداف تحقیق.....	۷
۱-۴-۱. هدف کلی تحقیق.....	۷
۲-۴-۱. اهداف ویژه تحقیق.....	۷
۳-۴-۱. اهداف کاربردی تحقیق.....	۸
۴-۴-۱. بهروران از نتایج تحقیق.....	۸
۵-۱. سوالات تحقیق.....	۸
۱-۵-۱. سوال اصلی تحقیق.....	۹
۲-۵-۱. سوالات فرعی تحقیق.....	۹
۶-۱. فرضیات تحقیق.....	۹
۷-۱. روش شناسی تحقیق.....	۱۰
۱-۷-۱. شرح روش تحقیق.....	۱۰
۲-۷-۱. شرح روش و ابزار گردآوری داده‌ها.....	۱۱
۳-۷-۱. جامعه آماری، روش نمونه گیری و حجم نمونه.....	۱۱
۴-۷-۱. روش‌ها و ابزار تجزیه و تحلیل داده‌ها.....	۱۲
۸-۱. تعریف واژه‌ها و اصطلاحات فنی و تخصصی.....	۱۲
۱-۸-۱. قراردادهای.....	۱۳
۹-۱. جمع بندی.....	۱۴

## فصل ۲ مروری بر ادبیات و پیشینه تحقیق

- ۱-۲. مقدمه ..... ۱۶
- ۲-۲. اصول و مفاهیم ریسک بانکی ..... ۱۷
  - ۱-۲-۲. تعریف ریسک از منظر بانکداری ..... ۱۷
  - ۲-۲-۲. انواع ریسک در بانکداری ..... ۱۸
  - ۳-۲-۲. مدیریت ریسک ساختار و نحوه اجرای آن در بانک ..... ۲۱
  - ۴-۲-۲. ریسک اعتباری، مفاهیم و تعاریف ..... ۲۲
- ۳-۲. امتیازدهی اعتباری ..... ۲۲
  - ۱-۳-۲. مروری بر تاریخچه ارزیابی ریسک اعتباری ..... ۲۲
  - ۲-۳-۲. تعریف امتیازدهی اعتباری ..... ۲۴
  - ۳-۳-۲. ضرورت و اهداف امتیازدهی اعتباری ..... ۲۵
  - ۴-۳-۲. مزایا و معایب امتیازدهی اعتباری ..... ۲۷
  - ۱-۴-۳-۲. منافع بالقوه امتیازدهی اعتباری ..... ۲۸
  - ۲-۴-۳-۲. معایب سیستم امتیازدهی اعتباری ..... ۲۹
  - ۵-۳-۲. اطلاعات مورد نیاز برای برآورد مدل امتیازدهی اعتباری ..... ۳۰
  - ۱-۵-۳-۲. معیار C ..... ۳۰
  - ۲-۵-۳-۲. معیار LAPP ..... ۳۲
  - ۳-۵-۳-۲. معیار P ..... ۳۲
  - ۶-۳-۲. مدل‌های امتیازدهی اعتباری ..... ۳۳
- ۴-۲. داده کاوی و کشف دانش از پایگاه داده ..... ۳۵
  - ۱-۴-۲. مفهوم داده کاوی ..... ۳۵
  - ۲-۴-۲. فرآیند کلی داده کاوی ..... ۳۶
  - ۳-۴-۲. روش‌های داده کاوی ..... ۳۸
- ۵-۲. طبقه بندی ..... ۴۰
  - ۱-۵-۲. الگوریتم‌های مختلف طبقه بندی ..... ۴۱



۴۱	۲-۵-۱-۱. طبقه بندی و امتیازدهی اعتباری
۴۲	۲-۵-۲. یادگیری ماشین و طبقه بندی
۴۲	۲-۶. درخت تصمیم
۴۴	۲-۶-۱. استخراج قاعده از یک درخت تصمیم
۴۴	۲-۶-۲. درخت تصمیم سنتی و انواع الگوریتم‌های آن
۴۵	۲-۶-۳. نقاط قوت و ضعف الگوریتم‌های درخت تصمیم
۴۶	۲-۶-۴. مروری بر الگوریتم‌های طبقه بندی مورد استفاده در تحقیق
۴۶	۲-۶-۴-۱. ساختار الگوریتمی درخت تصمیم
۴۸	۲-۶-۴-۲. معیار انتخاب مشخصه در الگوریتم‌های درخت تصمیم
۵۰	۲-۶-۴-۳. الگوریتم CART
۵۱	۲-۶-۴-۴. الگوریتم CHAID
۵۱	۲-۶-۴-۵. الگوریتم QUEST
۵۱	۲-۶-۴-۶. شبکه‌های بیزین
۵۲	۲-۶-۴-۷. شبکه‌های عصبی
۵۶	۲-۶-۴-۸. ماشین بردار پشتیبان
۵۷	۲-۶-۴-۹. رگرسیون لجستیک
۵۹	۲-۶-۴-۱۰. تحلیل تمایزی
۶۰	۲-۷. تصمیم‌گیری چند معیاره
۶۰	۲-۷-۱. تصمیم‌گیری
۶۱	۲-۷-۲. طبقه بندی فنون تصمیم‌گیری بر اساس معیار تصمیم‌گیری
۶۲	۲-۷-۲-۱. تصمیم‌گیری چند شاخصه
۶۳	۲-۷-۳. تصمیم‌گیری فنون MADM
۶۳	۲-۷-۳-۱. مدل سازی در فنون تصمیم‌گیری چند شاخصه
۶۸	۲-۷-۴. انواع روش‌های حل فنون تصمیم‌گیری چند شاخصه
۷۰	۲-۷-۵. روش TOPSIS

۷۱	..... فرآیند تحلیل سلسله مراتبی
۷۲	..... ۱-۶-۷-۲. محاسبه وزن‌های معیارها با استفاده از فرآیند تحلیل سلسله مراتبی
۷۳	..... ۲-۶-۷-۲. محاسبه نرخ سازگاری و ناسازگاری تصمیم
۷۳	..... ۸-۲. مروری بر تحقیقات پیشین
۷۹	..... ۹-۲. جمع بندی

### فصل ۳ متدولوژی تحقیق

۸۱	..... ۱-۳. مقدمه
۸۱	..... ۲-۳. فرآیند داده کاوی
۸۱	..... ۱-۲-۳. مدل فرآیندی داده کاوی بر اساس استاندارد CRISP_DM
۸۴	..... ۳-۳. داده‌های تحقیق
۸۴	..... ۱-۳-۳. روش‌های جمع آوری اطلاعات
۸۶	..... ۲-۳-۳. جامعه و نمونه آماری
۸۷	..... ۳-۳-۳. نوع داده‌ها و مقیاس آن‌ها
۸۷	..... ۴-۳. ساختار اجرایی تحقیق
۸۷	..... ۱-۴-۳. درک مسئله کسب و کار
۸۸	..... ۲-۴-۳. درک داده‌ها
۸۹	..... ۳-۴-۳. آماده سازی داده‌ها
۹۲	..... ۴-۴-۳. مدل سازی داده‌ها
۹۴	..... ۵-۴-۳. ارزیابی نتایج
۹۵	..... ۶-۴-۳. بکارگیری مدل
۹۵	..... ۵-۳. جمع بندی

### فصل ۴ تجزیه و تحلیل داده‌ها

۹۷	..... ۱-۴. مقدمه
۹۷	..... ۲-۴. توصیف داده‌ها

- ۹۷ ..... ۱-۲-۴. ویژگی‌های داده‌ها
- ۹۷ ..... ۱-۱-۲-۴. جامعه آماری
- ۹۷ ..... ۲-۱-۲-۴. نمونه‌های مورد بررسی پایگاه داده
- ۹۸ ..... ۳-۱-۲-۴. رکوردهای پایگاه داده
- ۹۸ ..... ۴-۱-۲-۴. فیلدهای پایگاه داده
- ۱۰۰ ..... ۵-۱-۲-۴. نوع و مقیاس داده‌ها
- ۱۰۲ ..... ۶-۱-۲-۴. ویژگی‌های خبرگان (کارشناسان بانک مورد مطالعه)
- ۱۰۲ ..... ۲-۲-۴. آماده سازی داده‌ها برای مدل
- ۱۰۲ ..... ۱-۲-۲-۴. درک مسئله کسب و کار
- ۱۰۳ ..... ۲-۲-۲-۴. درک داده‌ها
- ۱۰۴ ..... ۳-۲-۲-۴. پالایش داده‌ها
- ۱۰۴ ..... ۴-۲-۲-۴. کاهش داده‌ها
- ۱۰۵ ..... ۵-۲-۲-۴. تغییرات داده‌ها
- ۱۰۶ ..... ۳-۲-۴. برچسب گذاری طبقات مشتریان با استفاده از تصمیم گیری چند معیاره
- ۱۰۹ ..... ۳-۴. تحلیل داده‌ها
- ۱۰۹ ..... ۱-۳-۴. مدل سازی
- ۱۱۱ ..... ۱-۱-۳-۴. دقت الگوریتم
- ۱۱۲ ..... ۲-۱-۳-۴. اهمیت متغیرها
- ۱۱۳ ..... ۲-۳-۴. ارزیابی اعتبار و کارایی مدل
- ۱۱۳ ..... ۱-۲-۳-۴. ارزیابی اعتبار مدل
- ۱۱۳ ..... ۲-۲-۳-۴. کارایی مدل در مقایسه با الگوریتم‌های مختلف طبقه بندی
- ۱۲۰ ..... ۳-۲-۳-۴. نتایج ترکیب مدل‌های طبقه بندی
- ۱۲۱ ..... ۴-۴. جمع بندی

## فصل ۵ نتیجه گیری و پیشنهادات

۱۲۳	.....	مقدمه	۱-۵
۱۲۳	.....	نتایج و یافته‌های تحقیق	۲-۵
۱۲۴	.....	پیشنهادات	۳-۵
۱۲۵	.....	پیشنهادات برای تحقیقات آتی	۱-۳-۵
۱۲۵	.....	پیشنهادات برای نظام بانکی	۲-۳-۵
۱۲۷	.....	منابع	
۱۳۲	.....	واژه نامه	
۱۳۷	.....	علائم اختصاری	
۱۳۹	.....	چکیده انگلیسی	

## فهرست اشکال

صفحه	عنوان
	<b>فصل ۱ کلیات تحقیق</b>
	<b>فصل ۲ مروری بر ادبیات و پیشینه تحقیق</b>
۳۸	شکل ۱-۲. گام‌های اصلی فرآیند اکتشاف دانش از پایگاه داده
۴۲	شکل ۲-۲. ساختار کلی یک مدل امتیازدهی
۵۳	شکل ۳-۲. نمونه‌ای از شبکه عصبی
۵۷	شکل ۴-۲. ماشین بردار پشتیبان در حالت خطی
۶۹	شکل ۵-۲. فنون MADM براساس پردازش اطلاعات بر مبنای شاخص‌ها
	<b>فصل ۳ متدولوژی تحقیق</b>
۸۲	شکل ۱-۳. مراحل مدل فرآیندی داده کاوی بر اساس استاندارد CRISP-DM
	<b>فصل ۴ تجزیه و تحلیل داده‌ها</b>
۱۰۳	شکل ۱-۴. فرآیند انتخاب روش داده کاوی مورد تقاضای مسئله
۱۰۴	شکل ۲-۴. فرآیند پالایش اولیه و یکپارچه سازی پایگاه داده‌ها
۱۰۵	شکل ۳-۴. حذف برخی رکوردها
۱۰۵	شکل ۴-۴. حذف برخی فیلدها
۱۰۹	شکل ۵-۴. فراخوانی داده‌ها در نرم افزار
۱۱۰	شکل ۶-۴. استریم مدل سازی الگوریتم‌های طبقه بندی
۱۱۱	شکل ۷-۴. درخت تصمیم ایجاد شده
۱۱۳	شکل ۸-۴. اهمیت شاخص‌ها
۱۱۹	شکل ۹-۴. اولویت بندی الگوریتم‌ها به لحاظ میزان صحت
۱۱۹	شکل ۱۰-۴. استریم پیش بینی وضعیت اعتباری مشتریان جدید

شکل ۴-۱۱. استریم مدل‌های ترکیبی الگوریتم‌های طبقه بندی ..... ۱۲۰

صفحه	عنوان
	<b>فصل ۱ کلیات تحقیق</b>
	<b>فصل ۲ مروری بر ادبیات و پیشینه تحقیق.</b>
۲۸	جدول ۱-۲. منافع سیستم امتیازدهی برای مشتریان و بانکها
۶۴	جدول ۲-۲. ماتریس تصمیم
۷۲	جدول ۳-۲. تعیین ارجحیت گزینه‌ها در مقایسات زوجی
	<b>فصل ۳ متدولوژی تحقیق</b>
۸۴	جدول ۱-۳. وظایف کلی مدل استاندارد CRISP-DM
	<b>فصل ۴ تجزیه و تحلیل داده‌ها</b>
۹۸	جدول ۱-۴. چارچوب کلی پایگاه داده مورد بررسی
۹۹	جدول ۲-۴. عناوین فیلدها
۱۰۰	جدول ۳-۴. توزیع داده‌ها در هر فیلد
۱۰۱	جدول ۴-۴. حالات متغیرهای اسمی
۱۰۱	جدول ۵-۴. مقیاس‌های داده‌های عددی
۱۰۲	جدول ۶-۴. ویژگی‌های خبرگان
۱۰۶	جدول ۷-۴. یک شکل نمودن داده‌ها
۱۰۸	جدول ۸-۴. ضریب نزدیکی
۱۱۲	جدول ۹-۴. درخت تصمیم CART
۱۱۴	جدول ۱۰-۴. درخت تصمیم C۵
۱۱۵	جدول ۱۱-۴. درخت تصمیم CHAID

۱۱۵	جدول ۴-۱۲. درخت تصمیم QUEST
۱۱۶	جدول ۴-۱۳. تحلیل تمایزی
۱۱۶	جدول ۴-۱۴. ماشین بردار پشتیبان
۱۱۷	جدول ۴-۱۵. رگرسیون لجستیک
۱۱۷	جدول ۴-۱۶. شبکه‌های بیزین BayesNet
۱۱۸	جدول ۴-۱۷. شبکه‌های عصبی Neural Network
۱۱۸	جدول ۴-۱۸. مقایسه تطبیقی صحت الگوریتم‌ها
۱۲۱	جدول ۴-۱۹. مقایسه نتایج صحت در دو حالت مدل ترکیبی



# فصل ۱

## کلیات تحقیق

## ۱-۱. مقدمه

در طبقه بندی ریسک‌هایی که بانک با آن مواجه می‌شود، "ریسک اعتباری"<sup>۱</sup> جایگاه ویژه‌ای دارد؛ زیرا اولین نقش بانک در بازارهای مالی، جمع آوری سپرده‌های مردم و اعطای این سپرده‌ها به صورت وام است تا از طریق اعطای وام کسب درآمد نماید. طبیعی است که اگر بانک در شناسایی مشتریان و اعطای وام به آن‌ها موثر عمل ننماید، زیان هنگفتی را متوجه خود خواهد کرد. از این رو لازم است بانک‌ها از میان متقاضیان وام، تنها تعدادی را انتخاب نمایند که از ادای دین آن‌ها در مدت زمان مقرر مطمئن هستند.

مطابق اطلاعات اداره آمار و داده‌های بانک مرکزی جمهوری اسلامی ایران، آمار مطالبات معوق بانک‌ها از سال ۱۳۸۰ تا سال ۱۳۸۵ همواره رو به رشد بوده است. بر اساس آمار موجود در سال ۱۳۸۵ مجموع مطالبات سررسید گذشته و معوق بانک‌ها و مؤسسات اعتباری غیر بانکی، معادل ۲۱۷.۸۲۳ میلیارد ریال بوده است که در مقایسه با مانده تسهیلات اعطایی بانک‌ها و مؤسسات اعتباری غیر بانکی به بخش غیر دولتی که معادل ۱.۰۲۳.۰۲۹ میلیارد ریال است، تقریباً ۲۱٪ را تشکیل می‌دهد. مطالبات سررسید گذشته و معوق بانک‌ها و مؤسسات اعتباری غیر بانکی در سال ۱۳۸۵ نسبت به سال ۱۳۸۴، ۵۶٪ رشد داشته است، در حالیکه مانده تسهیلات بانک‌ها و مؤسسات اعتباری غیر بانکی به بخش غیر دولتی در همین دوره ۴۰٪ رشد نموده است [۱۴].

طبق آمار بانک مرکزی جمهوری اسلامی ایران، نسبت مطالبات معوق (شامل مطالبات از بخش دولتی و غیردولتی) به کل تسهیلات ریالی بانک‌ها و مؤسسات اعتباری در سال ۱۳۸۶، ۴.۱۱ درصد و در سال ۱۳۸۷ به ۵.۱۷ درصد رسیده است که به علت سیاست‌های دولت و شرایط نامناسب اقتصادی و رکود و تورم در کشور افزایش یافته است.

از طرفی این نسبت با ۴.۴ واحد درصد کاهش از ۳.۱۸ در سال ۱۳۸۸ به ۱۳.۹ درصد در سال ۱۳۸۹ کاهش یافت. طبق گزارشات رسمی رئیس کل بانک مرکزی، مطالبات معوق سیستم بانکی کشور را ۵۳۸ هزار میلیارد ریال تا سال ۱۳۹۰ اعلام شده است [۹].

سیستم بانکی در ایران نیز مانند سایر کشورها نقش بسیار مهمی در اقتصاد ایجاد می‌نماید. زیرا علاوه بر آنکه بانک‌ها واسطه وجوه در بازار پول هستند، به دلیل توسعه ناکافی بازار سرمایه، نقش اساسی

<sup>۱</sup> Credit Risk

در تامین مالی برنامه‌های میان مدت و بلند مدت اقتصادی کشور دارند. در کشور ما از یک سو به دلیل اندک بودن تشکیل سرمایه توسط بخش خصوصی و از سوی دیگر به دلیل عدم توسعه بازار سرمایه، امکان تشکیل سرمایه مورد نیاز در بخش‌های مختلف اقتصادی به سادگی وجود ندارد و بدیهی است که در این صورت، "اعطای تسهیلات"، بخش مهمی از عملیات در بانک را تشکیل دهد. در واقع بانک‌ها می‌توانند با عملیات اعتباری خود و ایجاد سرمایه مالی برای بخش‌های مختلف اقتصادی، شرایط مناسبی برای سرمایه گذاری فراهم آورند. در مجموع می‌توان مهم‌ترین فعالیت بانک‌ها را، جمع آوری منابع مالی و تخصیص آن‌ها به بخش‌های مختلف اقتصادی دانست [۱۴].

به منظور پوشش و مدیریت مخاطرات اعطای تسهیلات، لازم است رویه‌ها و مکانیزم‌های کنترلی به گونه‌ای ایجاد گردند که بتوان از تخصیص تسهیلات به متقاضیان، اطمینان حاصل کرد. ارزیابی صحیح پروژه‌ها و طرح‌ها نه تنها از آن جهت اهمیت دارد که طرح‌ها منبع سودآوری و بقای فعالیت بانک‌ها هستند، بلکه اهمیت این موضوع با توجه به محدود بودن سرمایه و لزوم هدایت سرمایه موجود در مسیر صحیح بیشتر مشخص می‌شود.

در گذشته روشی که برای شناسایی مشتریان (اعتبار سنجی) مورد استفاده قرار می‌گرفت، روش قضاوتی بود که تنها تصمیم شخص ارزیاب در راستای سیاست‌های مصوب بانک و موسسه اعتباری، در اعطای وام موثر بود و این مسئله موجب می‌شود در بعضی مواقع فرد اعتبار دهنده سلايق شخصی خود را در مسئله اعطای اعتبار دخیل نماید؛ که این روند رضایت مشتریان را کاهش می‌دهد. علاوه بر آن به دلیل اشتباه افراد، ناکافی بودن تجربه برخی از ارزیابان و ثبت نشدن سوابق و تجربیات گذشته، روش مناسبی برای اعتبار سنجی مشتریان به شمار نمی‌رود. امروزه روش‌هایی وجود دارد که می‌توان با استفاده از معیارهای مشخص، مشتریان را به دور از سلايق شخصی ارزیابی نمود. بانک‌ها و محققین بسیاری تلاش کرده‌اند با شناسایی این معیارها و تدوین مدل‌های مختلف، وام دهنده را در تصمیم‌گیری خود یاری نمایند [۱۴].

## ۲-۱. بیان مسئله تحقیق

یکی از موضوعات مهم در بانک‌ها و مؤسسات اعتباری، اتخاذ تصمیم صحیح جهت اعطای تسهیلات به متقاضیان وام می‌باشد. بانک‌ها همواره سعی می‌کنند به نحوی این وظیفه خود را انجام دهند،

تا احتمال عدم برگشت اصل و سود تسهیلات را کاهش دهند. به بیان دیگر بانکها به دنبال کاهش ریسک اعتباری می باشند. یکی از شیوه های مرسوم جهت کاهش ریسک اعتباری، رتبه بندی اعتباری متقاضیان وام است. بانکها با بهره گیری از کارشناسان خبره خود به این امر مبادرت می ورزند. اما به علت مسائلی همچون هزینه بالا، زمان بر بودن و کمبود کارشناسان خبره، با محدودیت هایی در این زمینه مواجه می باشند. ضمن آنکه در این شیوه سنجش اعتبار، قضاوت های شخصی دخیل می باشند. با گسترش فناوری های نوین اطلاعاتی و کاربردهای وسیع آن در پردازش و طبقه بندی اطلاعات، شیوه های جدیدی جهت تصمیم گیری، معرفی شده است. از جمله این شیوه های کاربردی، فرآیند داده کاوی اطلاعات مشتریان می باشد. بدین وسیله با بهره گیری از تحلیل اطلاعات مربوط به مشتریان بانک می توان به رتبه بندی اعتباری متقاضیان وام و طبقه بندی آن ها به مشتریان خوش حساب و بد حساب، بدون قضاوت شخصی و بر اساس سیستم های هوشمند پرداخت. این شیوه کمک خواهد کرد تا بانکها مشتریان خود را بهتر و دقیق تر شناسایی و رتبه بندی اعتباری نموده و نسبت به گذشته، ریسک اعتباری خود را بیشتر کاهش داده و منابع مالی خود را به صورت بهینه تری به مشتریان خوش حساب اختصاص دهند و از قبل، مشتریان جدید را با توجه به ویژگی هایشان طبقه بندی نمایند. از طرفی بهره گیری از ابزارهای نوین هوش مصنوعی می تواند در ارتقا سیستم های پشتیبانی از تصمیم گیری بانکها در حوزه مشتریان اعتباری بسیار راهگشا باشد و بانک های داخلی نیز همگام با بانک های کشورهای توسعه یافته بتوانند از این ابزارها در جهت به روز شدن سیستم های اعتباری و مدیریت ارتباط با مشتریان بهره مند گردند. بر همین اساس این تحقیق به دنبال شناسایی الگوریتم طبقه بندی مناسب جهت رتبه بندی اعتباری مشتریان بانک مورد مطالعه از میان الگوریتم های داده کاوی منتخب می باشد. تا یک سیستم پیش بینی کننده برای وضعیت ریسک اعتباری مشتریان جدید ارائه دهد. البته قبل از این کار این تحقیق ابتدا یک فرآیند برچسب گذاری نوین برای هر رکورد از داده های مشتریان ارائه می دهد که در ادامه تشریح گردیده است. در مسائل طبقه بندی، مانند امتیازدهی اعتباری، روش بدین صورت می باشد که یک مجموعه داده داریم که متشکل از تعدادی رکورد و تعدادی فیلد می باشد و یکی از فیلدها به عنوان کلاس بوده که از پیش تعیین شده است. از طرفی در هر کسب و کار، برای تخصیص حالت های کلاس به هر رکورد، به عبارتی برچسب زدن به هر رکورد، دو حالت رخ می دهد: