





دانشگاه الزهراء (س)

دانشکده فنی و مهندسی

پایان نامه جهت اخذ درجه کارشناسی ارشد

رشته مدیریت فناوری اطلاعات

عنوان

کاربرد داده کاوی در مدیریت ارتباط با مشتری (مورد کاوی بانک ملی ایران)

استاد راهنما

دکتر مهدی سیف برقی

استاد مشاور

دکتر محمد علی صنیعی منفرد

دانشجو

بهاره سیف

اردیبهشت ماه ۱۳۹۰

کلیه دستاوردهای این تحقیق متعلق به دانشگاه الزهراء (س) است.

تقدیم به

پدر و مادر عزیزم به پاس زحمات و فداکاریشان

و تقدیم به

همسر مهربانم به پاس دلگرمی هایش

که اگر همراهی آنان نبود این مهم به انجام نمی رسید

از زحمات اساتید گرانقدرم جناب آقای دکتر سیف برقی و  
جناب آقای دکتر صنیعی منفرد کمال تشکر و سپاسگذاری

را دارم

## چکیده

امروزه با پیشرفت فناوری اطلاعات، سیستم های نوینی ایجاد شده است که می تواند توان سازمان را در کاهش هزینه های داخلی، تعامل بهتر با محیط و در نهایت به دست آوردن سود یاری رساند. یکی از این ابزارها، مدیریت ارتباط با مشتری است. مدیریت ارتباط با مشتری نقش مهمی در شناخت نیازهای مشتریان و جهت دهی فعالیت های مختلف واحدهای صنعتی و خدماتی به منظور رفع نیاز مشتریان دارد. به علاوه، بهبود و توسعه فناوری اطلاعات امکان جمع آوری، ذخیره سازی و تحلیل حجم وسیعی از داده ها را با استفاده از تکنیک های داده کاوی فراهم نموده است. امروزه در بسیاری از جنبه های مدیریت ارتباط با مشتری، تحلیلگران به دنبال بررسی تکنیک های داده کاوی به منظور شناسایی رفتار مشتریان، الگوها و عملکردهای آنها جهت پیش بینی رفتار مشتریان و به تبع آن برآوردن تمایلات آنها بر اساس این الگوها هستند.

در این پژوهش ضمن مروری بر ادبیات داده کاوی و مدیریت ارتباط با مشتری و همچنین استفاده کاربردی این ابزار در آن، به منظور طبقه بندی مشتریان و همچنین پیش بینی طبقه مشتریان بانک ملی ایران مدلی بر اساس تکنیک درخت تصمیم C5 و با استفاده از داده های مشتریان سابق بانک مورد مطالعه ایجاد گردیده است. علاوه بر سنجش اعتبار خود مدل، مقایسه تطبیقی این تکنیک با برخی از مهمترین تکنیک های کلاسه بندی همچون درخت تصمیم CART، درخت تصمیم CHAID، درخت تصمیم QUEST، ماشین بردار پشتیبان، شبکه بیزین و رگرسیون لجستیک به منظور بررسی دقت و صحت مدل صورت گرفته است. نتایج حاصل از مقایسه تطبیقی صحت کلاسه بندها نشان داد که درخت تصمیم C5 بالاترین صحت را در تشخیص کلاس مشتریان دارد.

**کلمات کلیدی:** فناوری اطلاعات، مدیریت ارتباط با مشتری، کلاسه بندی، پیش بینی، درخت تصمیم

## فهرست مطالب

### فصل اول - کلیات تحقیق

۱-۱- مقدمه.....	۲
۲-۱- زمینه تحقیق.....	۳
۳-۱- دامنه تحقیق.....	۳
۴-۱- اهمیت و ضرورت انجام تحقیق.....	۳
۵-۱- اهداف تحقیق.....	۵
۱-۵-۱- اهداف کلی تحقیق.....	۵
۲-۵-۱- اهداف جزئی تحقیق.....	۵
۳-۵-۱- اهداف کاربردی تحقیق.....	۵
۶-۱- بهره وران از نتایج تحقیق.....	۶
۷-۱- سؤالات تحقیق.....	۶
۸-۱- فرضیه‌های تحقیق.....	۶
۹-۱- مدل اجرایی تحقیق.....	۶
۱۰-۱- شرح روش تحقیق.....	۸
۱۱-۱- روش‌ها و ابزار تجزیه و تحلیل داده‌ها.....	۹
۱۲-۱- محدودیت‌های تحقیق.....	۱۰

### فصل دوم: آشنایی با موضوع و مرور ادبیات

۱-۲- مقدمه.....	۱۲
-----------------	----

- ۲-۲- مدیریت ارتباط با مشتری ..... ۱۳
- ۱-۲-۲- مفهوم مدیریت ارتباط با مشتری ..... ۱۳
- ۲-۲-۲- تعریف جامعی از مدیریت ارتباط با مشتریان ..... ۱۵
- ۳-۲-۲- ابعاد مدیریت ارتباط با مشتری ..... ۱۶
- ۴-۲-۲- تعریف مفاهیم مطرح در مدیریت ارتباط با مشتری ..... ۱۶
- ۵-۲-۲- اهداف مدیریت ارتباط با مشتری ..... ۱۷
- ۳-۲- داده کاوی ..... ۱۸
- ۱-۳-۲- تعریف داده کاوی ..... ۱۹
- ۲-۳-۲- کشف دانش در پایگاه داده ها (KDD) ..... ۲۰
- ۱-۲-۳-۲- فرآیند کشف دانش ..... ۲۰
- ۳-۳-۲- روشهای داده کاوی ..... ۲۲
- ۱-۳-۳-۲- دسته بندی یا کلاسه بندی ..... ۲۴
- ۲-۳-۳-۲- پیش بینی ..... ۲۶
- ۴-۳-۲- لزوم داده کاوی ..... ۲۶
- ۴-۲- کاربرد داده کاوی در مدیریت ارتباط با مشتری ..... ۲۷
- ۱-۴-۲- داده کاوی در مدیریت ارتباط با مشتریان و بانکداری ..... ۲۷
- ۱-۱-۴-۲- ساختار سیستماتیک مدیریت ارتباط با مشتری بانک های تجاری ..... ۲۷
- ۲-۱-۴-۲- کاربرد داده کاوی در سیستم CRM بانک های تجاری ..... ۲۸



- ۳۰-۲-۴-۱-۳- مدیریت منابع مشتری بر مبنای داده کاوی.....
- ۳۱-۲-۴-۱-۴- جریان کاری داده کاوی در مدیریت ارتباط با مشتریان بانک های تجاری و آنالیز ریسک مشتری.....
- ۳۲-۲-۴-۱-۵- دسته بندی مشتری بانک و روش های طبقه بندی در مدیریت ریسک.....
- ۳۳-۲-۴-۱-۶- بهبود ارتباط بین بانک ها و مشتریان.....
- ۳۴-۲-۴-۲- اصول و مفاهیم ریسک بانکی.....
- ۳۴-۲-۴-۱- تعریف ریسک از منظر بانکداری.....
- ۳۴-۲-۴-۲- انواع ریسک در بانکداری.....
- ۳۵-۲-۴-۳- مدیریت ریسک، ساختار ونحوه اجرای آن در بانک.....
- ۳۵-۲-۴-۴- ریسک اعتباری؛ مفاهیم و تعاریف.....
- ۳۵-۲-۴-۳- امتیازدهی اعتباری.....
- ۳۶-۲-۴-۳-۱- تعریف امتیاز دهی اعتباری.....
- ۳۶-۲-۴-۳-۲- ضرورت و اهداف امتیاز دهی اعتباری.....
- ۳۹-۲-۴-۳- مزایا و معایب امتیازدهی اعتباری.....
- ۴۰-۲-۴-۳- مدل های امتیاز دهی اعتباری.....
- ۴۱-۲-۵-۵- مروری بر تحقیقات پیشین.....
- ۴۱-۲-۵-۱- مطالعات داخلی.....
- ۴۶-۲-۵-۲- مطالعات خارجی.....
- ۶۵-۲-۶- جمع بندی فصل.....

## فصل سوم : متدولوژی تحقیق

۶۷	۱-۳- مقدمه.....
۶۷	۲-۳- فرایند داده کاوی.....
۶۷	۱-۲-۳- مدل فرایندی داده کاوی بر اساس استاندارد CRISP-DM.....
۷۰	۳-۳- داده های تحقیق.....
۷۱	۱-۳-۳- شیوه و ابزار جمع آوری داده ها.....
۷۱	۲-۳-۳- نوع داده ها و مقیاس آنها.....
۷۱	۳-۳-۳- جامعه آماری و نمونه گیری.....
۷۲	۴-۳- اعتبار و کارایی روش ارائه شده.....
۷۳	۵-۳- ساختار اجرایی تحقیق.....
۷۳	۱-۵-۳- درک مساله کسب و کار.....
۷۴	۲-۵-۳- درک داده ها.....
۷۵	۳-۵-۳- آماده سازی داده ها.....
۷۸	۴-۵-۳- مدلسازی داده ها.....
۷۹	۵-۵-۳- ارزیابی نتایج.....
۷۹	۶-۵-۳- به کارگیری مدل.....
۸۰	۷-۵-۳- معرفی نرم افزار Spss Clementine.....
۸۱	۱-۷-۵-۳- مدیریت Clementine.....

- ۸۲.....Clementine های پروژه های ۲-۷-۵-۳
- ۸۲.....Streams جریان ها ۳-۷-۵-۳
- ۸۲.....معرفی الگوریتم های اعمال شده ۶-۳
- ۸۲.....درخت تصمیم ۱-۶-۳
- ۸۴.....استخراج قاعده از یک درخت تصمیم ۱-۱-۶-۳
- ۸۴.....درخت تصمیم سنتی و انواع الگوریتم های آن ۲-۱-۶-۳
- ۸۵.....ساختار الگوریتمی درخت تصمیم ۳-۱-۶-۳
- ۸۷.....معیار انتخاب مشخصه در الگوریتم درخت تصمیم ۴-۱-۶-۳
- ۸۹.....شبکه های بیزین ۲-۶-۳
- ۹۰.....ماشین بردار پشتیبان ۳-۶-۳
- ۹۴.....رگرسیون لجستیک ۴-۶-۳
- ۹۵.....تحلیل تمایزی ۵-۶-۳
- ۹۶.....جمع بندی ۷-۳

### فصل چهارم : تجزیه و تحلیل داده ها

- ۹۸.....مقدمه ۱-۴
- ۹۸.....توصیف داده ها ۲-۴
- ۹۸.....ویژگی های داده ها ۱-۲-۴
- ۹۸.....جامعه آماری ۱-۱-۲-۴

- ۹۸.....۲-۱-۲-۴ نمونه های مورد بررسی پایگاه داده
- ۹۹.....۳-۱-۲-۴ رکورد های پایگاه داده
- ۹۹.....۴-۱-۲-۴ فیلد های پایگاه داده
- ۱۰۰.....۵-۱-۲-۴ نوع و مقیاس داده ها
- ۱۰۲.....۶-۱-۲-۴ ویژگی های خبرگان(کارشناسان بانک مورد مطالعه)
- ۱۰۳.....۲-۲-۴ آماده سازی داده ها برای مدل
- ۱۰۳.....۱-۲-۲-۴ درک مساله کسب و کار
- ۱۰۴.....۲-۲-۲-۴ درک داده ها
- ۱۰۵.....۳-۲-۲-۴ پالایش داده ها
- ۱۰۵.....۴-۲-۲-۴ کاهش داده
- ۱۰۶.....۵-۲-۲-۴ تغییرات داده ها
- ۱۰۸.....۳-۴ تحلیل داده ها
- ۱۰۸.....۱-۳-۴ مدل سازی
- ۱۱۱.....۱-۱-۳-۴ صحت الگوریتم
- ۱۱۲.....۲-۱-۳-۴ اهمیت متغیرها
- ۱۱۳.....۲-۳-۴ ارزیابی اعتبار و کارایی مدل
- ۱۱۳.....۲-۲-۳-۴ کارایی مدل در مقایسه با الگوریتم های مختلف طبقه بندی
- ۱۲۷.....۴-۴ پیش بینی رکوردهای جدید ( پیش بینی کلاس مشتریان جدید بر اساس قوی ترین الگوریتم)

۴-۵- جمع بندی فصل ..... ۱۲۸

### فصل پنجم : نتیجه گیری و پیشنهادات

۵-۱- مقدمه..... ۱۳۰

۵-۲- نتایج و یافته های تحقیق ..... ۱۳۰

۵-۲-۱- پاسخ به سوال تحقیق و بررسی فرضیه..... ۱۳۰

۵-۲-۲- نتایج مستقیم حاصل از اجرای تحقیق..... ۱۳۱

۵-۲-۳- نتایج حاصل از مطالعات حوزه ادبیات تحقیق..... ۱۳۲

۵-۳- پیشنهادات..... ۱۳۲

۵-۳-۱- پیشنهادات برای تحقیقات آتی..... ۱۳۲

۵-۳-۲- پیشنهادات برای نظام بانکی..... ۱۳۳

منابع ..... ۱۳۵

ضمائم..... ۱۴۳

## فهرست اشکال

۸	شکل ۱-۲- مدل اجرایی تحقیق
۱۶	شکل ۱-۲- تعریف CRM
۲۰	شکل ۲-۲- فرآیند کشف دانش
۲۲	شکل ۳-۲- روشهای داده کاوی
۲۹	شکل ۴-۲- چارچوب دسته بندی تکنیک های داده کاوی در مدیریت ارتباط با مشتری
۴۹	شکل ۵-۲- چرخه داده کاوی - مدیریت ارتباط با مشتری
۵۴	شکل ۶-۲- مدل فرآیندهای بخش بندی مشتریان بر مبنای داده کاوی
۸۰	شکل ۲-۳- صفحه اصلی نرم افزار CLEMENTINE
۹۱	شکل ۴-۳- ماشین بردار پشتیبان در حالت خطی
۱۰۴	شکل ۱-۴- فرآیند انتخاب تکنیک داده کاوی مورد تقاضای مسئله
۱۰۵	شکل ۲-۴- فرایند پالایش اولیه و یکپارچه سازی پایگاه داده ها
۱۰۵	شکل ۳-۴- حذف برخی رکوردها
۱۰۶	شکل ۴-۴- حذف برخی فیلدها
۱۰۹	شکل ۵-۴- مدل درخت تصمیم C5
۱۱۳	شکل ۶-۴- اهمیت شاخص ها
۱۱۴	شکل ۷-۴- درخت تصمیم CART
۱۱۵	شکل ۸-۴- اهمیت شاخص ها در الگوریتم CART
۱۱۶	شکل ۹-۴- درخت تصمیم CHAID
۱۱۷	شکل ۱۰-۴- اهمیت شاخص ها در الگوریتم CHAID
۱۱۸	شکل ۱۱-۴- درخت تصمیم QUEST
۱۱۹	شکل ۱۲-۴- اهمیت شاخص ها در الگوریتم QUEST
۱۲۰	شکل ۱۳-۴- الگوریتم تحلیل تمایزی
۱۲۱	شکل ۱۴-۴- اهمیت شاخص ها در الگوریتم تحلیل تمایزی
۱۲۲	شکل ۱۵-۴- الگوریتم ماشین بردار پشتیبان
۱۲۳	شکل ۱۶-۴- الگوریتم رگرسیون لجستیک
۱۲۴	شکل ۱۷-۴- الگوریتم شبکه های بیزین
۱۲۶	شکل ۱۸-۴- اهمیت شاخص ها در الگوریتم شبکه های بیزین

- شکل ۴-۱۹- اولویت بندی الگوریتم ها به لحاظ میزان صحت (بر حسب درصد) ..... ۱۲۷
- شکل ۴-۲۰- جریان ایجاد شده به منظور پیش بینی کلاس احتمالی مشتری جدید ..... ۱۲۷
- شکل ۴-۲۱- نتایج پیش بینی کلاس احتمالی مشتریان جدید ..... ۱۲۸

## فهرست جداول

جدول ۱-۲	منافع سیستم امتیازدهی برای مشتریان و بانک ها	۴۰
جدول ۲-۲	رده بندی مقالات ارائه شده در زمینه کاربرد داده کاوی در فرآیند جذب مشتری بر اساس مدل و تکنیک مورد استفاده و مرجع آن	۵۸
جدول ۳-۲	دسته بندی مقالات ارائه شده در زمینه کاربرد DM در فرآیند شناسایی مشتری بر اساس مدل و تکنیک مورد استفاده	۵۹
جدول ۴-۲	مقالات ارائه شده در زمینه کاربرد DM در فرآیند نگهداری مشتری بر اساس مدل و تکنیک مورد استفاده و مرجع آن	۶۰
جدول ۱-۳	وظایف کلی مدل استاندارد CRISP-DM	۷۰
جدول ۱-۴	چارچوب کلی پایگاه داده مورد بررسی	۹۹
جدول ۲-۴	عناوین فیلدهای موجود در پایگاه داده	۱۰۰
جدول ۳-۴	شقوق (حالت‌های ممکن) متغیرهای اسمی	۱۰۱
جدول ۴-۴	مقیاس‌های داده‌های عددی	۱۰۲
جدول ۵-۴	ویژگی‌های خبرگان	۱۰۳
جدول ۶-۴	یک شکل نمودن داده‌ها	۱۰۷
جدول ۷-۴	طیف طبقات	۱۰۷
جدول ۸-۴	برخی از قوانین کشف شده توسط مدل	۱۱۰
جدول ۹-۴	درخت تصمیم C5	۱۱۲
جدول ۱۰-۴	ماتریس متقاطع مربوطه	۱۱۲
جدول ۱۱-۴	درخت تصمیم CART	۱۱۴
جدول ۱۲-۴	ماتریس متقاطع مربوطه	۱۱۵
جدول ۱۳-۴	درخت تصمیم CHAID	۱۱۶
جدول ۱۴-۴	ماتریس متقاطع مربوطه	۱۱۷
جدول ۱۵-۴	درخت تصمیم QUEST	۱۱۸
جدول ۱۷-۴	تحلیل تمایزی	۱۲۰
جدول ۱۸-۴	ماتریس متقاطع مربوطه	۱۲۱
جدول ۱۹-۴	ماشین بردار پشتیبان SUPPORT VECTOR MACHINE	۱۲۲
جدول ۲۰-۴	ماتریس متقاطع مربوطه	۱۲۳



۱۲۴	جدول ۴-۲۱- رگرسیون لجستیک LOGISTIC-R
۱۲۴	جدول ۴-۲۲- ماتریس متقاطع مربوطه
۱۲۵	جدول ۴-۲۳- شبکه های بیزین BAYESNET
۱۲۵	جدول ۴-۲۴- ماتریس متقاطع مربوطه
۱۲۶	جدول ۴-۲۵- مقایسه تطبیقی صحت الگوریتم ها

فصل اول

کلیات تحقیق

## ۱-۱- مقدمه

شرکتهای امروزی از طریق تجزیه و تحلیل چرخه زندگی مشتری به افزایش ارزش مشتری دست یافته اند. ابزارها و فناوریهای انبار داده، داده کاوی و دیگر تکنیک های مدیریت ارتباط با مشتری، روشهایی هستند که فرصتهای جدیدی را برای تجارت فراهم کرده اند. در واقع دیدگاه محصول محوری جای خود را به مشتری محوری داده است. بنابراین، با جمع آوری داده های مربوط به مشتری و تصمیم گیری براساس الگوهای استخراج شده از روابط پنهان میان داده ها به وسیله ابزار داده کاوی، می توان به خواسته مشتری محوری خود جامه عمل پوشاند [۱].

در سالهای اخیر فرهنگ تجارت به پیشرفتهایی نایل گشته است. مطابق با آن، روابط اقتصادی مشتریان به طور بنیادی و اساسی در حال تغییر است. شرکتهای به منظور نظارت بر این گونه تغییرات نیازمند ارائه راه حلها هستند. ظهور و پیدایش اینترنت در تغییر جهت مرکز توجه بازاریابی نقش بسزایی داشته است. چنانچه اطلاعات بر خط<sup>۱</sup> بیشتر در دسترس قرار گیرد موجب آگاهی و هوشیاری بیشتر مشتریان می گردد. آنها در جریان آنچه آرایه و پیشنهاد می شود قرار می گیرند و تقاضای بهترین ها را دارند. برای از عهده برآمدن در چنین شرایطی باید سیستم هایی که بتواند به طور دقیق نسبت به مشتریان واکنش نشان دهد به کار رود. جمع آوری آمار مشتریان و داده های رفتاری آنها این هدف اصلی و دقیق را ممکن می سازد. این نوع هدف گیری به یک برنامه ریزی عالی هنگام ایجاد یک رقابت سخت و به مشخص کردن مشتریان بالقوه هنگام عرضه محصولات جدید کمک می کند [۱].

اغلب کسب و کارها به تصمیم گیری های استراتژیک و یا اتخاذ خط مشی های جدید برای خدمت رسانی بهتر به مشتریان نیاز دارند. به عنوان مثال فروشگاهها آرایش مغازه خود را برای ایجاد میل بیشتر به خرید مجدداً طراحی می کنند و یا خطوط هواپیمایی تسهیلات خاصی را برای مشتریان جهت پروازهای مکرر آنها در نظر می گیرند. این دو مثال به داده هایی در مورد رفتار مصرفی گذشته مشتریان برای تعیین الگوهایی به وسیله داده کاوی، نیاز دارد. براساس این الگوها تصمیمات لازم اتخاذ می شود. در واقع ابزار داده کاوی، داده را می گیرد و یک تصویر از واقعیت به شکل مدل می سازد، این مدل روابط موجود در داده ها را شرح می دهد [۲].

---

1 On line

در این پژوهش به دنبال مقایسه تکنیک های داده کاوی به منظور تشخیص موثرترین تکنیک، جهت یافتن الگوهایی در میان داده های مربوط به مشتریان بانک و کلاسه بندی آنها هستیم تا بتوان با استفاده از الگوهای استخراج شده از داده ها به پیش بینی وضعیت اعتباری مشتریان پرداخت و به بانکها در اعطای تسهیلات بانکی به مشتریان و کاهش مطالبات معوق کمک کرد.

## ۱-۲- زمینه تحقیق

توسعه به کارگیری سیستمهای اطلاعاتی در سازمانها و گسترش انتقال فرآیندهای سازمانی به سامانه های الکترونیکی و نیز وابستگی روزافزون مدیران و تصمیم سازی ها به اطلاعات و سیستمهای اطلاعاتی، مباحث نوینی نظیر یکپارچه سازی<sup>۱</sup>، داده کاوی<sup>۲</sup> و مخزن داده<sup>۳</sup> را مطرح ساخته است. این پژوهش در حوزه سیستم های اطلاعاتی است.

## ۱-۳- دامنه تحقیق

پژوهش حاضر بر روی اطلاعات گرفته شده از تعدادی شعب بانک ملی ایران شهر تهران انجام شده است.

## ۱-۴- اهمیت و ضرورت انجام تحقیق

افزایش روز افزون حجم «مطالبات معوق» بانکها در ایران یکی از تازه ترین مظاهر نابسامانی های عمیقی است که نظام بانکی را طی دورانی طولانی، از ایفای وظایف اصلی خود در پیشبرد توسعه کشور باز داشته است. مأموریت اصلی نظام بانکی، ایجاد ارتباط میان پس اندازکنندگان و آن دسته از بازیگران عرصه اقتصادی است که به دریافت تسهیلات مالی نیاز دارند.

وجود «مطالبات معوق» به آن معناست که دریافت کنندگان تسهیلات مالی و یا به عبارت دیگر، بدهکاران نظام بانکی یا از باز پرداخت اصل و فرع بدهی های خود فرار می کنند یا به دلایلی، از انجام این کار ناتوانند.

---

<sup>1</sup> Integration

<sup>2</sup> Data mining

<sup>3</sup> Data Warehouse