



بانک کشوری
بانک مردم ایران
الکترونیک

مرکز پژوهش‌های کاربردی

طرح تحقیق

پایان‌نامه دانشجویی

PROPOSAL

خلاصه اطلاعات طرح پروپوزال

عنوان پروپوزال به زبان فارسی:

بهبود نظام اعتبار سنجی مشتریان براساس رویکرد داده محوری

عنوان پروپوزال به زبان انگلیسی:

Improving the Customer Credit Scoring System Based on a Data-Driven Approach

کلمات کلیدی:

وام های بانکی، پیش بینی نکول، هوش مصنوعی، شبکه عصبی مصنوعی، پرسپترون چندلایه،
الگوریتم کلونی مورچه

مشخصات موضوعی پروپوزال

عنوان پروپوزال:

بهبود نظام اعتبار سنجی مشتریان براساس رویکرد داده محوری در بانک کشاورزی

طرح مسئله و ضرورت انجام تحقیق:

۱. عنوان مساله

در این پروپوزال، قصد داریم به بهبود سیستم‌های اعتبارسنجی داخلی مشتریان در بانک کشاورزی با استفاده از تکنولوژی هوش مصنوعی و فرآیند داده‌کاوی بپردازیم. هدف از این اقدام، کاهش ریسک بانک‌ها در اعطای تسهیلات است.

۲. ضرورت انجام پروژه

هر کسب و کار و تجارتی برای حفظ و توسعه بازار خود مجبور به اعطای اعتبارات کوتاه مدت، میان مدت و گاهاً بلند مدت به مشتریان می‌باشد. اعتبارات کوتاه مدت در این چارچوب معمولاً برای حجم کم اعتبار و خدمات معمول ارائه شده اعطا می‌گردند. تنوع ریسک در اعطای این اعتبارات به نحوی است که مدیریت آن معمولاً از طریق نظام‌های بازاریابی و یا حسابداری مؤسسات انجام می‌پذیرد، چون موضوع عدم پرداخت یا تأخیر در پرداخت معمولاً به سادگی قابل مدیریت می‌باشد. اما اعتبارات میان مدت و گاه بلند مدت برای مشتریان همراه با ریسک و عدم اطمینان است. احتمال عدم ایفای تعهدات مالی (نکول) در این ریسک‌ها میتواند در بازارهای مختلف مبادله کالا، خدمات، بازارهای سرمایه و یا لیزینگ ضرورت پرداختن به این موضوع را افزایش میدهد.

معوقات بانکی مشکل مؤسسات مالی در اکثر کشورهای جهان است. بررسی‌های بانک جهانی در سال 2020 حاکی از آن است که بی ثباتیهای اقتصادی، افزایش بیرویه وام دهی، مداخلات دولتها در امر وام دهی و کافی نبودن نظارت بر پرداخت تسهیلات از جمله دلایل اصلی پیدایش معوقات بانکی در سطح منطق‌های و جهانی است.

هر چهار مورد یاد شده در نظام بانکی ایران به وضوح قابل مشاهده می‌باشند. این عوامل در کنار تفاوت قابل توجه نرخ سود بانکی با نرخ واقعی باعث بروز ریسک اعتباری در بانکها شده است.

در بانک کشاورزی علاوه بر موارد فوق، ریسک‌های تولید در بخش کشاورزی که اغلب ناشی از شرایط آب و هوایی و حوادث غیرمترقبه می‌باشد، موجب افزایش ریسک اعتباری تسهیلات کشاورزی نیز شده است.

خلاصه اجرایی

در این پروژه توسعه قصد داریم یک سیستم نرم افزاری جامع و یکپارچه در فرآیند اعتبارسنجی مشتریان؛ طراحی و پیاده سازی کنیم. هدف این سیستم کاهش ریسک‌های اعتباری و افزایش دقت و کارایی در فرآیند اعتبارسنجی مشتریان است.

این سیستم با استفاده از تجزیه و تحلیل پیشرفته داده، یادگیری ماشین و شبکه‌های عصبی مصنوعی پرسپترون چند لایه و الگوریتم یادگیری پس انتشار ارتجاعی برای آموزش شبکه، بانک‌ها را قادر می‌سازد تا تصمیم‌گیری‌های مبتنی بر داده‌ها، کاهش ریسک در اعطای تسهیلات، افزایش دقت در ارزیابی اعتبار مشتریان، افزایش بهره‌وری، افزایش رضایت مشتریان داشته باشد.

گام	فعالیت اجرایی
1	<ul style="list-style-type: none"> • شناسایی وضعیت فعلی اعتبارسنجی مشتریان در بانک کشاورزی • شناسایی دقیق اهداف اعتبارسنجی مشتریان و گردآوری نیازها و انتظارات آنها به صورت دقیق • تعیین و تشریح منابع و اقلام داده‌های مورد نیاز • شناسایی نقاط ضعف وقوت موجود
2	<ul style="list-style-type: none"> • مدل سازی داده ای و استخراج الگوریتم های مناسب برای سنجش شاخص ها • شاخص سازی و طراحی معیارها و ابزارهای مناسب اعتبارسنجی مشتریان • طراحی الگوریتمهای مناسب برای پیش بینی ریسک اعتباری و عملکرد مشتریان

<ul style="list-style-type: none"> طراحی الگو پیشنهادهای هوشمند متناسب با رتبه اعتباری مشتری مبتنی بر طیف تصمیم گیری طراحی ساختار داده ها و دیتابیس های مورد نیاز برای ذخیره و پردازش اطلاعات 	
<ul style="list-style-type: none"> گردآوری داده های تاریخی مشتریان از سیستم های مختلف بانکی پاکسازی و تصفیه داده ها برای حذف نواقص و اطمینان از کیفیت داده ها یکپارچه سازی داده ها از منابع مختلف برای ایجاد پایگاه داده های جامع 	3
<ul style="list-style-type: none"> آموزش مدل های هوش مصنوعی با استفاده از داده های جمع آوری شده آزمایش و اعتبارسنجی مدلها برای اطمینان از دقت و کارایی آنها تنظیم و بهینه سازی مدلها براساس نتایج آزمایش ها 	4
<ul style="list-style-type: none"> توسعه رابط های برنامه نویسی کاربردی (API) برای اتصال مدل های اعتبارسنجی به سیستم های بانکی اطمینان از همگام سازی و هماهنگی بین سیستم های جدید و سیستم های قدیمی آزمایش و ارزیابی عملکرد سیستم های یکپارچه شده 	5

جدول مراحل بنیادی پروژه و فعالیت های اجرایی مربوطه

تاریخچه و شرح مسئله

در دهه های گذشته، نظام های اعتبارسنجی عمدتاً به روش های سنتی و مبتنی بر اطلاعات محدود و قضاوت های انسانی متکی بودند. این سیستم ها معمولاً از داده های ساختاریافته مانند گزارش های مالی، سوابق اعتباری، و اطلاعات جمعیتی برای ارزیابی اعتبار مشتریان استفاده می کردند. این روش ها به دلیل محدودیت در دسترسی به داده های جامع و پیشرفته، دقت و کارایی پایینی داشتند.

با پیشرفت فناوری اطلاعات و ظهور داده کاوی و هوش مصنوعی، امکان بهره گیری از داده های گسترده و الگوریتم های پیچیده برای ارزیابی دقیق تر و سریع تر اعتبار مشتریان فراهم شده است. این رویکرد داده محور، نه تنها

سرعت و دقت فرآیندهای اعتبارسنجی را افزایش می‌دهد، بلکه امکان شناسایی الگوهای پنهان و تحلیل جامع‌تری از رفتار مالی مشتریان را فراهم می‌کند.

در نظام‌های سنتی اعتبارسنجی، چندین چالش و مشکل وجود دارد که ضرورت بهبود آنها با استفاده از رویکرد داده‌محوری و روش‌های هوش مصنوعی را نمایان می‌سازد:

دقت پایین: روش‌های سنتی اغلب قادر به تحلیل دقیق تمامی جوانب رفتار مالی مشتریان نیستند و این می‌تواند منجر به ارزیابی‌های نادرست و افزایش ریسک اعتباری شود.

زمان‌بر بودن: فرآیندهای اعتبارسنجی سنتی معمولاً زمان‌بر هستند و نیازمند دخالت انسانی زیادی هستند که می‌تواند باعث تأخیر در تصمیم‌گیری‌ها و کاهش رضایت مشتریان شود.

عدم انعطاف‌پذیری: سیستم‌های سنتی اعتبارسنجی اغلب توانایی تطبیق سریع با تغییرات بازار و شرایط اقتصادی را ندارند و این می‌تواند باعث کاهش کارایی آنها شود.

محدودیت اطلاعاتی: استفاده از داده‌های محدود و ساختاریافته در روش‌های سنتی، توانایی تحلیل جامع و شناسایی الگوهای پنهان در رفتار مالی مشتریان را محدود می‌کند.

فرضیات تحقیق:

روش شبکه عصبی پرسپترون چندلایه با انتخاب ویژگی براساس الگوریتم کلونی مورچه عملکرد دقیق‌تری نسبت به روش شبکه عصبی پرسپترون چندلایه با تمام ویژگی‌ها، در پیش‌بینی نکول وام‌های بانکی و بهبود نظام اعتبارسنجی دارد.

اعتبارسنجی

اعتبار به معنی "اعتماد داشتن و باور کردن" مترادف کلمه credit است. اعتبار در مفهوم عام آن، حسن شهرتی است که شخص نسبت به قابلیت ایفای تعهداتش دارد، یعنی اعتماد به انجام تعهدات یک شخص در طول انجام یک معامله یا معاملاتی. بر این اساس، واژه اعتبار در عرف بنگاه‌های اقتصادی، مترادف با ایجاد تسهیلات مالی برای مشتریان، اعتبارات بانکی، تسهیلات اعتباری و تسهیلات اعطایی می‌باشد.

Thomas در سال 2009 اعتبارسنجی را این‌گونه تعریف کرد "فرایند شناخت مشتریان بانک به منظور اعطای اعتبار به آنها براساس یک مجموعه‌ای از معیارهای از پیش تعریف شده"

ریسک اعتباری

ریسک اعتباری عبارت است از احتمال عدم وصول تسهیلات اعطایی به اشخاص در نتیجه ورشکستگی یا تنزل موقعیت مالی و اعتباری شخص دریافت کننده تسهیلات که به لحاظ ماهیت آن در بخش اعطای تسهیلات موسسات مالی و بالاخص بانکها وجود دارد. به عبارت دیگر، ریسک اعتباری به این معنی است که باز پرداخت های مشتریان به بانک در قبال تسهیلات دریافتی ممکن است با تأخیر انجام گیرد و یا حتی اصلاً وصول نشود، که این امر به نوبه خود مسایل و مشکلاتی را برای جریان وجوه نقد و نیز مدیریت نقدینگی بانک به همراه خواهد آورد.

شبکه عصبی

یک شبکه عصبی مصنوعی، در واقع یک پارادیم محاسباتی ریاضی عمومی است که عملیات سیستمهای عصبی بیولوژیکی را مدل می نماید. تمام مدل های شبکه های عصبی که در طی سال ها ارائه شده اند، دارای بلوک مشترکی به نام نرون و یک ساختار شبکه ای متصل به هم هستند. مدل نرونی که بیشتر از همه مورد استفاده قرار گرفته بر پایه نرون مک کولاج و پیترز بنا نهاده شده و پرسترون چندلایه نام شبکه ای است که بیشتر از همه مورد استفاده قرار گرفته و بر پایه اتصال به ترتیب چندین لایه از پرسترون ها بنا نهاده شده است. شبکه های عصبی یکی از از متداول ترین فنون داده کاوی در طبقه بندی و پیشگویی در حوزه مالی بوده و میتواند برای پیشگویی هر سری زمانی مالی استفاده شود. یکی از مزایای شبکه های عصبی، کاربرد وسیع آن ها در زمینه های مختلف مسائل مالی بوده و ابزارهای که از شبکه عصبی حمایت می کنند، تحت هر محیطی یافت میشوند. بازنمایی مختلف داده ها در شبکه های عصبی می تواند نتایج مختلفی را تولید نماید. بنابراین تنظیم داده ها نقش مهمی در استفاده از آنها دارد. شبکه عصبی عبارت از یک سیستم ورودی-خروجی است که پردازش اطلاعات توسط لایه های پنهان انجام میگردد. بنابراین هر سیستم ورودی-خروجی با سازه شبکه عصبی دارای اجزای زیر است:

- لایه های ورودی: شامل بردار متغیرهای مستقل
- لایه های پنهان: شامل توابع عملگر، نرون های پردازش کننده داده های ورودی
- لایه خارجی: شامل برآورد (پیش بینی) بردار متغیر هدف (وابسته)

در ادبیات شبکه عصبی به جای اصطلاح تخمین ضرایب از اصطلاح یادگیری یا آموزش برای پیدا کردن ارزش های وزنها شبکه استفاده می شود. هدف کمینه کردن مجموع مربعات خطا یعنی اختلاف میان ارزشهای خروجی واقعی و خروجی برازش شده از شبکه عصبی می باشد. بدین منظور داده ها را به دو مجموعه آموزش و آزمون تقسیم می کنند، اولی برای الگو سازی و یادگیری و دیگری برای پیشبینی استفاده می شود.

الگوریتم بهینه سازی کلونی مورچه

الگوریتم کلونی مورچه الهام گرفته شده از مطالعات و مشاهدات روی کلونی مورچه هاست. یکی از مهمترین جالبترین رفتار مورچه ها، رفتار آنها برای یافتن غذا است، به ویژه چگونگی پیدا کردن کوتاهترین مسیر میان منابع غذایی و آشیانه. این نوع رفتار مورچه ها دارای نوعی هوشمندی تودهای است که اخیراً مورد استفاده قرار میگیرد.

چالش ها و ریسک ها

• چالش های فنی

- **کیفیت داده ها:** کیفیت پایین داده های ورودی می تواند نتایج مدل سازی و تحلیل ها را تحت تأثیر قرار دهد، که به اصلاح و پاکسازی داده ها نیاز دارد.
- **یکپارچه سازی سیستم ها:** ترکیب و یکپارچه سازی سیستم های مختلف و داده های متعدد میتواند چالش برانگیز باشد و نیاز به هماهنگی دقیق داشته باشد.
- **پیچیدگی الگوریتم ها:** پیاده سازی الگوریتم های پیچیده یادگیری ماشین و هوش مصنوعی نیازمند تخصص بالا و منابع محاسباتی قوی است.

• چالش های عملیاتی

- **مدیریت تغییرات:** انتقال از سیستم های سنتی به یک سیستم داده محور نیازمند تغییرات گسترده در فرآیندها و فرهنگ سازمانی است.
- **آموزش کارکنان:** کارکنان باید با سیستم های جدید و تکنولوژی های مورد استفاده آشنا شوند.

• چالش های مدیریتی و سازمانی

- **هماهنگی بین تیم ها:** هماهنگی بین تیم های مختلف با تخصص های متنوع نیازمند مدیریت قوی و توانایی برقراری ارتباط مؤثر است.
- **مدیریت منابع:** مدیریت بهینه منابع مالی، انسانی و تجهیزاتی به ویژه در پروژه های بزرگ و پیچیده اهمیت دارد.
- **تغییرات سازمانی:** تغییرات ناگهانی در ساختار سازمانی یا تیم مدیریتی می تواند بر پیشرفت پروژه تأثیر بگذارد.

• ریسک های مالی

- **هزینه‌های بالا:** هزینه‌های پیاده‌سازی، نگهداری و به‌روزرسانی سیستم‌های داده‌محور ممکن است بیشتر از پیش‌بینی‌ها باشد.
- **بازگشت سرمایه نامشخص:** بازگشت سرمایه ممکن است به اندازه کافی سریع و قابل توجه نباشد تا هزینه‌های پروژه را توجیه کند.

• ریسک‌های قانونی و مقرراتی

- **تغییرات در قوانین و مقررات:** تغییرات ناگهانی در قوانین و مقررات محلی، ملی یا بین‌المللی می‌تواند منجر به توقف یا تغییر در روند پروژه شود.
- **مجوزها و تاییدیه‌ها:** تأخیر در اخذ مجوزها و تاییدیه‌های لازم از سوی مراجع ذیصلاح می‌تواند زمانبندی پروژه را تحت تأثیر قرار دهد.

• ریسک‌های تکنولوژیکی

- **پیشرفتهای تکنولوژیکی:** پیشرفت‌های سریع در فناوری ممکن است نیاز به روزرسانی تجهیزات و سیستم‌های موجود را افزایش دهد.
- **امنیت سایبری:** حملات سایبری و نقض‌های امنیتی می‌تواند منجر به از دست دادن داده‌ها و اطلاعات حساس شود و روند پروژه را مختل کند.

• ریسک‌های انسانی

- **کمبود نیروی انسانی ماهر:** کمبود نیروی انسانی با مهارت‌های تخصصی می‌تواند پروژه را با تأخیر مواجه کند.
- **تعهد و انگیزه پرسنل:** کاهش انگیزه و تعهد پرسنل می‌تواند منجر به کاهش بهره‌وری و کیفیت کار شود.

اهداف کلی:

به طور کلی، هدف این پروژه بهبود نظام اعتبار سنجی مشتریان جهت پیش‌بینی نکول وام‌های بانکی که باعث کاهش میزان ریسک متقبل شده از طرف بانکها می‌شود، و شناسایی مشتریانی که احتمال بازپرداختن وام از سوی

آنها زیاد است میتواند هزینه سرمایه بانک ها و موسسات مالی را کاهش دهد، لذا در این تحقیق سعی میکنیم با ارائه یک مدل کارآمد این مشتریان را شناسایی کرده و هزینه سرمایه و ریسک بانکها را کاهش دهیم.

در حقیقت با بررسی مشخصات مشتریان قصد شناسایی مشتریان بالقوه بدحساب را داشته و در این راه دو روش ماشین بردار پشتیبان و شبکه عصبی پرسپترون چند لایه را با یکدیگر مقایسه می کنیم.

طرح پیشنهادی

همانطور که گفتیم طرح پیشنهادی شامل 9 مرحله بوده که در سه مرحله بنیادی تعبیه شده اند. در جدول زیر جزئیات هر یک از مراحل ارائه می شود:

مرحله		توصیف
اول	جذب داده	برای جذب داده ها از MySQL، PostgreSQL، Microsoft SQL Server، Oracle
دوم	ذخیره سازی داده	ذخیره داده های دریافت شده در یک قالب ساختاریافته با استفاده از PostgreSQL و MongoDB برای داده های بدون ساختار
سوم	پردازش داده	استفاده از Apache Spark برای پردازش داده در مقیاس بزرگ و کتابخانه pandas و Numpy برای کارهای پاکسازی داده های کوچکتر و درون حافظه
چهارم	پاک سازی داده ها	استفاده از OpenRefine برای پاک سازی داده ها و تبدیل فرمت ها و DataRobot برای پاک سازی و آماده سازی داده ها به همراه مدل سازی خودکار
پنجم	تجزیه و تحلیل داده ها	استفاده از مدل های یادگیری ماشینی با استفاده از Scikit-learn یا TensorFlow و Pytorch برای تجزیه و تحلیل داده ها
ششم	مصور سازی	ایجاد داشبوردهای تعاملی را با Plotly Dash یا Streamlit برای تجسم و گزارش
هفتم	اتوماسیون	حصول اطمینان از کل گردش کار با استفاده از Apache Airflow هماهنگ کنید، از خودکار بودن وظایف و مدیریت وابستگی ها

هشتم	تجزیه و تحلیل بلادرنگ	اجرای پردازش بلادرنگ داده با Apache Spark Streaming برای تجزیه و تحلیل داده ها در زمان رسیدن
نهم	استقرار در سامانه	استقرار برنامه در زیرساخت های مناسب برای اطمینان از مقیاس پذیری و قابلیت اطمینان

نقشه راه و پیاده سازی اولیه

در این بخش به بیان نقشه راه در قالب فازبندی و مراحل پروژه و سپس به ذکر مثال هایی از پیاده سازی اولیه می پردازیم. در انتها نیز به یک نمونه از پروژه عملی انجام شده در بانک های دیگر اشاره می کنیم. ذکر این مثال ها و نمونه پروژه انجام شده در بانک های دیگر صرفاً جهت آشنایی کلی با موضوع بهبود نظام اعتبار سنجی مشتریان بهره گیری از تکنیک های هوش مصنوعی، یادگیری ماشین و نمونه های کاربردهای آن بوده و بسته به تقاضای کارفرما، ماهیت و پیچیدگی داده ها و نیز الگوریتم های انتخابی بهینه یادگیری ماشین جهت مدلسازی و ارزیابی دقت و پیش بینی احتمالی، ممکن است دستخوش تغییراتی شود.

فازبندی و مراحل پروژه

طبق مراحل اصلی و فعالیت های اجرایی توصیف شده، نقشه راه پیشنهادی به تفصیل نشان داده شده است:

گام اجرایی		توصیف فعالیت طبق مراحل اجرایی
1	جمع آوری داده	جذب و جمع آوری داده ها ذخیره سازی و مدیریت داده ها
2	مدلسازی داده	پردازش و پاکسازی داده ها تجزیه و تحلیل داده ها و یادگیری ماشین مصورسازی و گزارش دهی داده ها
3	توسعه	تصمیم گیری در مورد ریسک های احتمالی طبق داشبورد مصورسازی و گزارش دهی

<p>اتوماسیون و ارکستراسیون گردش کار</p> <p>تجزیه و تحلیل داده های زمان واقعی</p> <p>طراحی سامانه، خدمات ابری و زیرساخت (بسته به تقاضای کارفرما)</p>		
---	--	--

جمع آوری و آماده سازی داده ها:

جمع آوری داده های مرتبط: شامل داده های تاریخی مشتریان، سوابق پرداخت، تراکنش ها، اطلاعات دموگرافیک، و داده های خارجی مثل اطلاعات اقتصادی و اعتباری.

پیش پردازش داده ها: شامل پاک سازی، رفع نواقص و ناهنجاری ها، نرمال سازی و تبدیل داده ها به فرمت های قابل استفاده

ابزارها: spark، pandas، dask

تحلیل داده ها و استخراج ویژگی ها

تحلیل توصیفی داده ها: تحلیل ابتدایی داده ها برای شناسایی الگوها، توزیع ها و همبستگی های موجود

استخراج ویژگی ها: شناسایی ویژگی های مهم و تاثیر گذار بر اعتبارسنجی، ایجاد و انتخاب ویژگی های جدید در صورت لزوم

ابزارها: Jupyter Notebook، PyTorch، TensorFlow، scikit-learn

ساخت مدل های اعتبارسنجی

انتخاب الگوریتم های مناسب: انتخاب الگوریتم های یادگیری ماشین مثل رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، شبکه های عصبی، کلونی مورچه و غیره.

آموزش و ارزیابی مدل ها: آموزش مدل ها با استفاده از داده های تاریخی و ارزیابی عملکرد آن ها با استفاده از معیارهایی مثل دقت، حساسیت، ویژگی، و F1-Score

ارزیابی و بهینه‌سازی مدل‌ها

مقایسه مدل‌ها: ارزیابی و مقایسه مدل‌های مختلف براساس عملکرد آن‌ها

بهینه‌سازی مدل‌ها: تنظیم پارامترها و اعمال تکنیک‌های بهینه‌سازی برای بهبود دقت و کارایی مدل‌ها

پیاده‌سازی و استقرار

پیاده‌سازی مدل نهایی: استقرار مدل منتخب در سیستم‌های عملیاتی سازمان

ایجاد زیرساخت‌های لازم: ایجاد زیرساخت‌های فنی برای اجرای مدل‌ها و ادغام آن‌ها با سیستم‌های موجود

پایش و به‌روزرسانی

پایش مستمر: نظارت بر عملکرد مدل‌ها به صورت دوره‌ای و شناسایی انحرافات

به‌روزرسانی مدل‌ها: به‌روزرسانی مدل‌ها براساس تغییرات در داده‌ها، رفتار مشتریان و شرایط بازار

مستندسازی و گزارش‌دهی

مستندسازی فرآیندها و نتایج: ثبت تمامی مراحل پروژه، روش‌ها، مدل‌ها و نتایج به دست آمده

گزارش‌دهی به ذینفعان: تهیه گزارش‌های دوره‌ای برای ذینفعان سازمانی در مورد عملکرد و بهبودهای حاصل از سیستم اعتبارسنجی جدید

ابزارها: Plotly Dash، Streamlit، Tableau، Matplotlib، Seaborn

توسعه

▪ ارزیابی ریسک‌های احتمالی:

- ابزارها: گزارشات داشبورد از مرحله قبلی جهت پیش بینی
- هدف: تحلیل ریسک‌های احتمالی مرتبط با سامانه اعتبارسنجی

```

1 import pandas as pd
2 import requests
3
4 # مثال: خواندن داده‌ها از یک فایل CSV
5 customer_data = pd.read_csv('customer_data.csv')
6
7 # مثال: دریافت داده‌ها از یک API
8 response = requests.get('https://api.example.com/customers')
9 api_data = response.json()
10 customer_data_api = pd.DataFrame(api_data)
11

```

▪ اتوماسیون و ارکستراسیون گردش کار:

• ابزارها: Luigi، Prefect، Apache Airflow

• هدف: اتوماسیون اعتبارسنجی و اطمینان از اجرای روان جریان های کاری پیچیده.

▪ طراحی سامانه، خدمات ابری و زیرساخت:

• ابزارها: در صورت تمایل کارفرما به خدمات ابری، میتوان از زیرساخت های ابری پیشنهادی توسط ابراروان، آسیاتک، پارس پک یا ایران سرور استفاده نمود.

• هدف: زیرساخت مقیاس پذیر برای ذخیره سازی داده ها، پردازش و نیز استقرار برنامه ها.

مثال هایی از پیاده سازی

در ادامه به بیان نمونه مثال هایی از پیاده سازی اولیه مراحل مختلف پروژه جهت آشنایی کلی می پردازیم:

نمونه کدهای اجرای اولیه پروژه

جمع آوری و مدیریت داده ها

نیازها: جمع آوری، ذخیره سازی و مدیریت کارآمد داده ها

این بخش شامل دریافت داده‌ها از منابع مختلف مانند فایل‌های CSV، دیتابیس‌ها یا API ها است.

پیاده سازی: استفاده از pandas برای دستکاری و پردازش داده ها و SQLAlchemy برای مدیریت

پایگاه داده

پاک سازی داده ها

در این بخش داده‌های ناقص یا نادرست شناسایی و اصلاح می‌شوند

```

1 import matplotlib.pyplot as plt
2
3 # بررسی توزیع امتیاز اعتباری
4 plt.hist(customer_data_cleaned['credit_score'], bins=20)
5 plt.title('توزیع امتیاز اعتباری مشتریان')
6 plt.xlabel('امتیاز اعتباری')
7 plt.ylabel('تعداد')
8 plt.show()
9

```

ویژگی سازی داده‌ها:

ایجاد ویژگی‌های جدید از داده‌های موجود

```

1 # حذف رکوردهای دارای مقادیر Null
2 customer_data_cleaned = customer_data.dropna()
3
4 # جایگزینی مقادیر نادرست
5 customer_data_cleaned['credit_score'] = customer_data_cleaned['credit_score'].replace([-1], customer_data_cleaned['credit_score'].mean())
6

```

تحلیل داده‌ها:

تحلیل‌های اولیه برای شناخت بهتر داده‌ها

```

1 # ایجاد ویژگی‌های جدید مانند نسبت بدهی به درآمد
2 customer_data_cleaned['debt_to_income_ratio'] = customer_data_cleaned['debt'] / customer_data_cleaned['income']
3

```

مدل سازی داده‌ها:

ایجاد مدل‌های ماشین یادگیری برای پیش‌بینی امتیاز اعتباری یا دسته‌بندی مشتریان

ارائه و گزارش گیری :

ایجاد داشبوردها و گزارش ها برای ارائه نتایج به ذینفعان

```
1 from sklearn.model_selection import train_test_split
2 from sklearn.ensemble import RandomForestClassifier
3 from sklearn.metrics import accuracy_score
4
5 # تقسیم داده ها به بخش های آموزشی و آزمون
6 X = customer_data_cleaned.drop('credit_score', axis=1)
7 y = customer_data_cleaned['credit_score']
8 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
9
10 # ایجاد مدل
11 model = RandomForestClassifier(n_estimators=100, random_state=42)
12 model.fit(X_train, y_train)
13
14 # پیش بینی و ارزیابی مدل
15 y_pred = model.predict(X_test)
16 accuracy = accuracy_score(y_test, y_pred)
17 print(f'دقت مدل: {accuracy}')
```

روش تحقیق جهت دستیابی به هریک از اهداف تحقیق، بیان جامعه آماری، نمونه و روش نمونه گیری، روش گردآوری اطلاعات و الگوی تجزیه و تحلیل مورد استفاده (مدل آماری یا اقتصادسنجی)

```
1 import seaborn as sns
2
3 # نمایش نمودارهای مهم
4 sns.pairplot(customer_data_cleaned[['income', 'debt', 'credit_score']])
5 plt.show()
6
```


نمونه پروژه انجام شده در زمینه بهبود اعتبارسنجی مشتریان توسط

هوش مصنوعی

از آنجایی که در پروژه قصد شناسایی مشتریانی است که احتمال نکول آنها بالاست، تحقیق از لحاظ هدف کاربردی است. این درحالی است که تحقیق بر روی داده های تاریخی صورت میپذیرد، از این رو میتوان آن را تحقیقی تجربی و از لحاظ متغیر، کمی میباشد.

جامعه آماری

این پروژه بر روی مشتریان حقیقی که از ۲۵ شعبه بانک مسکن واقع تهران اقدام به دریافت وام خرید مسکن (با سپرده) کرده اند، انجام شده است.

نمونه آماری و روش نمونه گیری

نمونه گیری به روش خوشه ای انجام پذیرفت و حجم نمونه بدلیل مشخص نبودن واریانس کل و میزان خطا از روش مورگان تعیین گردید. خوشه ها در تحقیق بانکهای شمال شرق شهرستان تهران فرض شدند که حجم نمونه گیری با توجه به نسبت مشتریان هر بانک به کل بانک های محدوده مشخص شده برآورد گردید.

تکنیکهای مورد استفاده

به منظور انجام پروژه پیشرو از روش شبکه عصبی پرسپترون چندلایه با انتخاب ویژگی برای پیشبینی نکول وامهای بانکی استفاده گردید. عملکرد این روش برای طبقه بندی مشتریان خوش حساب و بدحساب با عملکرد روش شبکه عصبی با در نظر گرفتن تمام ویژگی ها، مقایسه شد و نتایج حاصله از استفاده دو مدل با آزمون میانگین مقایسه زوجی مورد سنجش قرار گرفت.

بررسی سابقه مطالعات و تحقیقات در موضوع طرح در داخل و خارج و چگونگی ارتباط آن با تحقیقات قبلی (منابع و مأخذ ضمیمه شود)

مالهاترا بررسی ای بر روی رتبه بندی اعتباری بانک های آمریکا انجام داد و مدل شبکه عصبی فازی را با روش آنالیز تفکیک غیر خطی مقایسه کردند. در نهایت به این نتیجه رسیدند که روش شبکه عصبی فازی نتایج بسیار بهتری را نسبت به روش سنتی آنالیز تفکیک غیر خطی به ارمغان می آورد (Malhotra and Malhotra, 2002)

آنجلینی تحقیقی را برای ارزیابی ریسک اعتباری شرکتهای کوچک فعال در ایتالیا انجام داد. در این تحقیق، او روش استاندارد شبکه عصبی پیشخور را با شبکه عصبی ای با معماری خاص مقایسه کرد و به این نتیجه رسید که روش با معماری خاص مشتریانی که تمایل به نکول دارند را بهتر از روش دیگر تشخیص میدهد (Angelini et al, 2008)

تساوی در تحقیقی به پیش بینی ورشکستگی مالی و ریسک اعتباری با استفاده از گونه های متفاوت از مدل پرسپترون چندلایه پرداختند. آنها با استفاده از الگوریتم پس انتشار اقدام به آموزش مدل ها کردند و روش های تک طبقه بندی کننده ساده، چندگانه، چندگانه معکوس را با یکدیگر مقایسه کردند. با مقایسه این روش ها در سه گروه از داده ها، پرسپترون چندلایه با طبقه بندی چندگانه تنها در یک مورد از سه مورد بهتر از نوع ساده بود. نتایج حاصله از پرسپترون چندلایه با طبقه بندی چندگانه معکوس از دو مورد دیگر نیز بدتر بود (Tsai and Wu,2008)

نقش ها، مسئولیت ها و منابع انسانی:

نقش ها و مسئولیت ها:

مدیر پروژه (Project Manager)

- مدیریت پروژه تجربه در پروژه های بزرگ و پیچیده
- برنامه ریزی، بودجه بندی، هدایت و نظارت بر اجرای پروژه
- مهارت های ارتباطی و توانایی برقراری ارتباط موثر با تیم ها و ذینفعان
- توانایی شناسایی و حل مسائل و مشکلات به صورت کارآمد
- مدیریت زمان، بودجه و منابع پروژه

متخصص داده (Data Specialist)

- دانش علوم داده تخصص در تحلیل داده ها و توسعه مدل های هوش مصنوعی
- برنامه نویسی تسلط بر زبان های برنامه نویسی مانند Python، R
- یادگیری ماشین تجربه در توسعه و پیاده سازی مدل های یادگیری ماشین و یادگیری عمیق
- ابزار های تحلیل داده تسلط بر ابزار های تحلیل داده مانند TensorFlow، PyTorch

تحلیلگر داده (Data Analyst)

- تحلیلگر داده توانایی جمع آوری، پاکسازی و تحلیل داده ها
- توانایی تهیه گزارش های تحلیلی و تفسیر داده ها
- تسلط بر نرم افزار های تحلیل داده مانند Excel، SQL، Tableau
- مهارت های ارتباطی توانایی ارائه نتایج تحلیل به صورت موثر به تیم های مختل

مدیریت IT (IT Manager)

- مدیریت IT تجربه در مدیریت زیرساخت های IT و پشتیبانی از تیم های فنی
- آشنایی با مفاهیم و پروتکل های امنیت سایبری
- توانایی نگهداری و ارتقاء سیستم های IT
- تسلط بر نرم افزارهای مدیریت IT و ابزارهای شبکه

توسعه دهنده نرم افزار (Software Developer)

- توسعه نرم افزار تجربه در توسعه و پیاده سازی نرم افزارهای تخصصی
- برنامه نویسی تسلط بر زبانهای برنامه نویسی مانند «Python، Java، ++C»
- یکپارچه سازی سیستمها توانایی یکپارچه سازی سیستمها و داده ها
- تست و رفع اشکال توانایی تست و رفع اشکالات نرم افزاری

متخصص امنیت سایبری (Cybersecurity Specialist)

- امنیت سایبری تخصص در حفاظت از سیستمها و داده ها در برابر تهدیدات سایبری
- ارزیابی امنیتی توانایی ارزیابی امنیتی سیستمها و شناسایی نقاط ضعف
- مدیریت رخدادهای امنیتی تجربه در مدیریت رخدادهای امنیتی و پاسخ به تهدیدات
- پروتکلهای امنیتی آشنایی با پروتکلهای و استانداردهای امنیتی

زمان بندی و گانت چارت پروژه:

جدول زمانی			
زمان	تکمیل این پروژه ماه مطابق با جدول زمانی زیر تخمین زده شده است		
۳ ماه	<ul style="list-style-type: none"> ▪ جذب و جمع آوری داده ها ▪ ذخیره سازی و مدیریت داده ها 	جمع آوری داده	1
۳ ماه	<ul style="list-style-type: none"> ▪ پردازش و پاکسازی داده ها ▪ تجزیه و تحلیل داده ها و یادگیری ماشینی 	مدل سازی داده	2

	مصورسازی و گزارش دهی داده ها		
۳ ماه	<ul style="list-style-type: none"> ▪ تصمیم گیری در مورد ریسک های احتمالی طبق داشبورد مصورسازی و گزارش دهی ▪ اتوماسیون و ارکستراسیون گردش کار ▪ تجزیه و تحلیل داده های زمان واقعی • طراحی سامانه، خدمات ابری و زیرساخت 	توسعه	3

سایر ملاحظات:

ضمائم زیر پس از پذیرش پروپوزال ارائه خواهند شد:

- رزومه پرسنل کلیدی پروژه
- نامه های حمایت از ذینفعان
- نمودارهای معماری فنی دقیق
- طرح های اولیه مراحل اجرایی مختل
- نمودار پروژه گانت با نقاط عطف
- تفکیک بودجه با توجیه هزینه

پیرو مذاکرات انجام شده با کارفرما محترم، جناب آقای کی قبادی بعد از برگزاری جلسه B2B و تأیید اولیه پیشنهاد فنی ، پیشنهاد مالی تقدیم حضورتان خواهد شد .

در جدول زیر لیست پرسنل کلیدی پروژه نمایش داده شده است

ردیف	نام و نام خانوادگی	کد ملی	رشته و مدرک تحصیلی	شغل	مؤسسه متبوع	درصد همکاری	تلفن همراه
1	محمدپارسا سهرابی	1749458586	دکتری مدیریت	مدیر فنی و اجرایی	پتروپالاتوس	25	09123344295
2	پریسا زارع	4420391723	ارشد ریاضی محض	برنامه نویس هوش مصنوعی	پتروپالاتوس	25	09162152204
3	زهره عظیمی	0016684631	کارشناسی it	برنامه نویس هوش مصنوعی	پتروپالاتوس	25	09382951028
4	سجاد بنزاده	-	کامپیوتر گرایش هوش مصنوعی و رباتیک	برنامه نویس هوش مصنوعی	پتروپالاتوس	25	09331251531

منابع

منابع فارسی

شیرین بخش، شمس اله ، یوسفی، ندا ، قربان زاد، جهان گیر، بررسی عوامل مؤثر بر احتمال عدم بازپرداخت تسهیلات اعتباری بانکها (مطالعه موردی مشتریان حقوقی بانک توسعه صادرات ایران)دانش مالی تحلیل اوراق بهادار،)

اخباری، مهدیه، رتبه بندی اعتباری مشتریان حقوقی بانک ها با رویکرد هوش مصنوعی ، پایان نامه کارشناسی ارشد، دانشگاه صنعتی اصفهان

صفری، سعید ، ابراهیمی شقاقی، مرضیه ، شیخ، محمدجواد ، مدیریت ریسک اعتباری مشتریان حقوقی در بانک های تجاری با رویکرد تحلیل پوششی داده ها (رتبه بندی اعتباری)، پژوهش های مدیریت در ایران،

منابع انگلیسی

Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. The quarterly review of economics and finance,

Comitato di Basilea per la vigilanza bancaria. (2004). International convergence of capital measurement and capital standards: a revised framework. Bank for International Settlements.

Basel Committee. (2010). Basel III: A global regulatory framework for more resilient banks and banking systems. Basel Committee on Banking Supervision, Basel.

Hu, Y. H., & Hwang, J. N. (Eds.). (2001). Handbook of neural network signal processing. CRC press.

Khashman, A. (2011). Credit risk evaluation using neural networks: Emotional versus conventional models. Applied Soft Computing,

Khashman, A. (2010). Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes. Expert Systems with Applications,

Kutsurelis, J. E. (1998). Forecasting financial markets using neural networks: An analysis of methods and accuracy (Doctoral dissertation, Monterey, California. Naval Postgraduate School).

Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Systems with Applications,