

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده علوم مالی

پایان نامه کارشناسی ارشد رشته ریاضی مالی

# کاربرد شبکه‌های عصبی مصنوعی در یافتن داده‌های گمشده

استاد راهنما

استاد مشاور

دانشجو

تیر ۱۴۰۲

ج

## تقدیر و تشکر

اکنون که به یاری پروردگار و یاری و راهبانی اساتید بزرگ، موفق به پایان این رساله شده‌ام، وظیفه خود دانسته که نهایت سپاسگزاری را از تمامی عزیزانی که در این راه به من کمک کرده‌اند، به عمل آورم.

و خالصانه از تمامی اساتید، معلمان و مدرسانی که در مقطع مختلف تحصیلی به من علم آموخته و مرا از سرچشمه دانایی سیراب کرده‌اند، تشکر می‌کنم.

## چکیده

امروزه، داده‌ها یکی از ارکان اصلی برای فرایند تصمیم‌گیری در سازمان‌های مختلف شناخته می‌شوند. با افزایش حجم و پیچیدگی داده‌ها، مقابله با داده‌های گمشده به یکی از چالش‌های اساسی در حوزه تحلیل داده تبدیل شده است. روش‌ها و رویکردهای مختلفی برای حل این مسئله وجود دارد. در این پژوهش، ضمن تعریف و تفصیل داده‌های گمشده، برخی از مدل‌های موجود برای جانهی داده‌های گمشده مالی بیان شده است. مدل‌های بیان‌شده بر اساس میزان خطا مورد ارزیابی و مقایسه قرار داده شده‌اند. به‌طور کلی، روش‌های موجود را می‌توان در دو دسته روش‌های آماری و روش‌های مبتنی بر یادگیری ماشین و شبکه‌های عصبی مصنوعی دسته‌بندی کرد. در این پژوهش، موردهای مورد ارزیابی شامل روش‌های رگرسیون خطی، رگرسیون چندجمله‌ای، رگرسیون لاسو، رگرسیون جنگل تصادفی، رگرسیون تقویت گرادیان، رگرسیون پرسپترون چندلایه، شبکه عصبی بازگشتی، K-نزدیک‌ترین همسایه و روش میانگین متحرک خودهمبسته یکپارچه خودکار هستند که جهت بررسی کارایی و دقت روش‌های مورد ارزیابی، روی داده‌های تاریخی ۸ سهم بورس اوراق بهادار تهران پیاده‌سازی شده‌اند. همچنین، مقایسه عملکرد روش‌های ذکر شده بیان شده است.

**کلمات کلیدی:** داده‌های گمشده، یادگیری ماشین، شبکه‌های عصبی مصنوعی

## فهرست مطالب

صفحه	عنوان
۱	فصل اول: کلیات پژوهش .....
۲	۱-۱ مقدمه .....
۳	۲-۱ بیان مسئله (تعریف و اهمیت موضوع) .....
۷	۱-۲-۱ انواع شبکه‌های عصبی مصنوعی .....
۸	۳-۱ اهداف تحقیق .....
۹	۴-۱ نوآوری تحقیق .....
۹	۵-۱ تعریف واژگان کلیدی .....
۹	۱-۵-۱ شبکه‌های عصبی مصنوعی .....
۹	۲-۵-۱ یادگیری ماشین .....
۱۰	۳-۵-۱ داده‌های گمشده .....
۱۰	۶-۱ ساختار کلی تحقیق .....
۱۱	فصل دوم: داده‌های گمشده .....
۱۲	۱-۲ مقدمه .....
۱۳	۲-۲ تکامل روش‌های تخمین داده‌های گمشده .....
۱۴	۳-۲ نسبت داده‌های گمشده .....
۱۵	۴-۲ ساختارهای گمشدگی داده‌ها .....
۱۸	۵-۲ الگوهای داده‌های گمشده .....
۸۰	۶-۴ نتیجه‌گیری .....
۸۳	۷-۴ پیشنهادهای کاربردی .....
۸۴	۸-۴ عناوین پیشنهادی برای تحقیقات آتی .....
۸۵	منابع فارسی .....
۸۶	منابع لاتین .....

## فهرست جدول‌ها

- جدول ۱-۱ برخی از کاربردهای شبکه‌های عصبی مصنوعی در زمینه‌های مختلف ..... ۵
- جدول ۱-۲ الگوی تک‌متغیره داده‌های گمشده ..... ۱۹
- جدول ۲-۲ الگوی داده‌های گمشده دلخواه ..... ۱۹
- جدول ۳-۲ الگوی داده‌های از دست رفته یکنواخت ..... ۱۹
- جدول ۱-۴ میزان خطاهای محاسبه شده در رگرسیون خطی ..... ۵۸
- جدول ۲-۴ میزان خطاهای محاسبه شده در رگرسیون لاسو ..... ۶۱
- جدول ۳-۴ میزان خطاهای محاسبه شده در رگرسیون چند جمله‌ای ..... ۶۴
- جدول ۴-۴ میزان خطاهای محاسبه شده در رگرسیون جنگل تصادفی ..... ۶۷
- جدول ۵-۴ میزان خطاهای محاسبه شده در رگرسیون تقویت گرادیان ..... ۷۰
- جدول ۶-۴ میزان خطاهای محاسبه شده در رگرسیون پرسپترون چند لایه ..... ۷۳
- جدول ۷-۴ خطای MAE بر آورده شده در مدل‌ها ..... ۸۲

## فهرست تصاویر

- شکل ۱-۱ انواع شبکه‌های عصبی همراه با شماتیک هر یک ..... ۷
- شکل ۱-۳ نمایش یک شبکه باور عمیق ..... ۳۵
- شکل ۲-۳ یک شبکه عصبی بازگشتی ..... ۴۵
- شکل ۱-۴ ساختار پژوهش ..... ۵۵
- شکل ۲-۴ نمودارهای رگرسیون خطی (مقدار واقعی-مقدار پیش‌بینی شده) ..... ۵۶
- شکل ۳-۴ هیستوگرام‌های اختلاف مقدار واقعی و مقدار پیش‌بینی شده توسط رگرسیون خطی ..... ۵۷
- شکل ۴-۴ نمودارهای رگرسیون لاسو (مقدار واقعی-مقدار پیش‌بینی شده) ..... ۵۹
- شکل ۵-۴ هیستوگرام‌های اختلاف مقدار واقعی و مقدار پیش‌بینی شده توسط رگرسیون لاسو ..... ۶۰
- شکل ۶-۴ نمودارهای رگرسیون چندجمله‌ای (مقدار واقعی-مقدار پیش‌بینی شده) ..... ۶۲
- شکل ۷-۴ هیستوگرام‌های اختلاف مقدار واقعی و مقدار پیش‌بینی شده توسط رگرسیون چندجمله‌ای ..... ۶۳
- شکل ۸-۴ نمودارهای رگرسیون جنگل تصادفی (مقدار واقعی-مقدار پیش‌بینی شده) ..... ۶۵
- شکل ۹-۴ هیستوگرام‌های اختلاف مقدار واقعی و مقدار پیش‌بینی شده توسط رگرسیون جنگل تصادفی ..... ۶۶
- شکل ۱۰-۴ نمودارهای رگرسیون تقویت گرادیان (مقدار واقعی-مقدار پیش‌بینی شده) ..... ۶۸
- شکل ۱۱-۴ هیستوگرام‌های اختلاف مقدار واقعی و مقدار پیش‌بینی شده توسط رگرسیون تقویت گرادیان ... ۶۹
- شکل ۱۲-۴ نمودارهای رگرسیون پرسپترون چندلایه (مقدار واقعی-مقدار پیش‌بینی شده) ..... ۷۱
- شکل ۱۳-۴ هیستوگرام‌های اختلاف مقدار واقعی و مقدار پیش‌بینی شده توسط پرسپترون چندلایه ..... ۷۲
- شکل ۱۴-۴ نمودارهای خطی مقدار جانیهشده توسط شبکه عصبی بازگشتی و مقدار واقعی ..... ۷۵
- شکل ۱۵-۴ نمودارهای خطی مقدار جانیهشده توسط روش K-نزدیک‌ترین همسایه و مقدار واقعی .. ۷۷
- شکل ۱۶-۴ نمودارهای خطی مقدار جانیهشده توسط روش Auto-ARIMA و مقدار واقعی ..... ۸۰
- شکل ۱۷-۴ رتبه‌بندی مدل‌های جانیهی بر اساس مقدار میانگین خطای MAE ..... ۸۳

## فصل اول: کلیات پژوهش



## ۱-۱ مقدمه

در دنیای داده‌محور امروز، اهمیت داده‌ها و تحلیل آن‌ها را نمی‌توان نادیده گرفت. هر زمینه‌ای، از امور مالی و مراقبت‌های پزشکی گرفته تا حمل و نقل و بازاریابی، به شدت وابسته به قدرت داده برای هدایت تصمیم‌گیری و به دست آوردن دیدگاه‌های ارزشمند است. از این رو، در دسترس بودن داده‌ها و کیفیت آن‌ها برای رسیدن به استراتژی‌های مؤثر از اهمیت بالایی برخوردار است.

داده‌ها نقش یک ستون را برای تعریف عملیات مختلف، بهینه‌سازی و بهبود عملکرد و ایجاد نوآوری در سازمان‌های مختلف ایفا می‌کنند. می‌توان گفت در سازمان‌های امروزی، داده تبدیل به یک ارز ارزشمند شده است که توسط آن بتوانند گوی رقابت را از رقبای خود برابند. با افزایش دیجیتال‌سازی اطلاعات، حجم و پیچیدگی داده‌های تولیدشده به سطوح بی‌سابقه‌ای رسیده است. از این رو، استفاده از قدرت نهفته در داده‌ها و یافتن الگوهای پنهان در آن‌ها برای رسیدن به تصمیم‌گیری‌های بهتر برای بقاء سازمان‌های مختلف، امری حیاتی است.

تجزیه و تحلیل داده‌ها و تفسیر سیستماتیک آن‌ها به عنوان یک ابزار ضروری برای استخراج دانش و استراتژی‌های اجرایی از داده‌ها عمل می‌کنند که شامل طیف وسیعی از تکنیک‌ها و روش‌های طراحی شده برای تجزیه و تحلیل مجموعه داده‌های بزرگ و متنوع، کشف الگوها، روندها و روابطی است که می‌تواند در تصمیم‌گیری‌های مختلف حائز اهمیت باشد.

با کنترل قدرت تحلیل داده‌ها، سازمان‌ها می‌توانند به مزیت رقابتی دست یابند. در حوزه تجارت، تجزیه و تحلیل داده‌ها شرکت‌ها را قادر می‌سازد تا رفتار مشتری را درک کنند، روندهای بازار را شناسایی کنند و پیش‌بینی‌هایی را بر اساس داده‌ها انجام دهند. این کار به شکل‌دهی کمپین‌های بازاریابی، بهینه‌سازی زنجیره‌های تأمین و هدایت سرمایه‌گذاری‌های استراتژیک کمک می‌کند و در نهایت، سودآوری و رضایت مشتری را افزایش می‌دهد.

در حوزه بهداشت و درمان، تجزیه و تحلیل داده‌ها نقش مهمی در تحقیقات بالینی، مراقبت از بیمار و مدیریت سلامت عمومی ایفا می‌کند. با تحلیل داده‌های بیمار، متخصصان پزشکی می‌توانند عوامل خطرناک را شناسایی کنند، برنامه‌های درمان شخصی را توسعه دهند و در کل، کیفیت خروجی نظام سلامت را بهبود ببخشند. علاوه بر این، سازمان‌های بهداشت عمومی از تحلیل داده برای پیگیری شیوع بیماری، پیش‌بینی اپیدمی و تخصیص مؤثر منابع استفاده می‌کنند و رفاه جامعه را در زمینه سلامت تضمین می‌کنند.

در امور مالی، تجزیه و تحلیل داده‌ها به مؤسسات مالی قدرت می‌دهد تا تصمیمات سرمایه‌گذاری آگاهانه بگیرند، ریسک‌ها را مدیریت کنند و فعالیت‌های کلاهبردانه را کشف کنند. با تجزیه و تحلیل داده‌های بازار، الگوها و روندهای تاریخی، تحلیلگران مالی می‌توانند رفتار بازار را پیش‌بینی کنند، پرتفوی را بهینه کنند و استراتژی‌های تجاری پیچیده‌ای را وضع کنند.

تأثیر تجزیه و تحلیل داده‌ها تنها به این زمینه‌ها محدود نمی‌شود. از حمل و نقل و مدیریت انرژی گرفته تا رسانه‌های اجتماعی و سرگرمی، تجزیه و تحلیل داده‌ها در اکثر صنایع، در حال تغییر رویکرد آن‌ها و رو آوردن به نوآوری در زمینه‌های مربوطه است.

با این حال، در میان دریای گسترده داده‌ها، اغلب یک چالش به وجود می‌آید: داده‌های گمشده. داده‌های گمشده می‌توانند به دلیل عوامل مختلفی مانند خطای انسانی، خطاهای سیستمی و غیره رخ دهند. حضور داده‌های گمشده می‌تواند مانع اثربخشی و دقت تجزیه و تحلیل داده‌ها شود که منجر به نتایجی با سوگیری و تفاسیر ناقص می‌شود.

برای غلبه بر این مشکل، شبکه‌های عصبی مصنوعی به عنوان یک راه‌حل قدرتمند ظاهر شده‌اند. این مدل‌های محاسباتی، با الهام از شبکه‌های عصبی مغز انسان، توانایی یادگیری از الگوهای موجود در داده‌ها و پیش‌بینی یا جانمایی اطلاعات از دست رفته، با دقت قابل توجهی را دارند.

در این مطالعه، ما کاربردهای شبکه‌های عصبی مصنوعی را در پرداختن به مشکل داده‌های گمشده بررسی خواهیم کرد. ما تکنیک‌ها و روش‌های به کار گرفته شده توسط این شبکه‌ها را بررسی خواهیم کرد، پتانسیل آن‌ها را برای افزایش دقت تحلیل داده‌ها و باز کردن دریچه‌های جدید در زمینه‌های مختلف نشان خواهیم داد. با بهره‌برداری از قابلیت‌های شبکه‌های عصبی مصنوعی، سازمان‌ها می‌توانند شکاف‌های موجود در مجموعه داده‌های خود را پر کنند و دیدگاه‌های قوی‌تری را به دست آورند و در نهایت، تصمیم‌های هوشمندانه‌تری بگیرند.

## ۱-۲ بیان مسئله (تعریف و اهمیت موضوع)

در دنیای امروز، اهمیت داده‌ها و تحلیل آن‌ها انکارناپذیر است. داده‌ها می‌توانند نقاط قوت و نقاط ضعف یک سیستم را برایمان آشکار سازند. داده‌ها می‌توانند علاوه بر تحلیل‌های گذشته‌نگر، ما را در تحلیل‌های آینده‌نگر نیز کمک کنند. استفاده از یادگیری ماشینی، هوش مصنوعی و ابزارها و روش‌های مرتبط با آن در حال افزایش است و همه این‌ها در بهبود سرعت و دقت تحلیل‌های ما