

## عنوان: مطالعه مقایسه‌ای روش‌های لجستیک و یادگیری ماشین در پیش‌بینی مدیریت سود

- بازه زمانی پژوهش رو یک دوره ۱۲ ساله از سال ۱۳۸۹ تا ۱۴۰۰ تشکیل و تعداد نمونه ۱۸۰ شرکت می‌باشد.
- تعداد متغیرهای مستقل موضوع پژوهش ۶۴ مورد

### ابزار تجزیه و تحلیل

پس از انتخاب متغیرهای بهینه پیش‌بین و پیش‌بینی مدیریت سود با استفاده از روش‌های لجستیک و یادگیری ماشینی مختلف به وسیله نرم‌افزار Weka نسخه ۳-۹، Rapidminer و سایر نرم افزارهایی که می‌توان روش یادگیری مورد نیاز را انجام داد، به منظور آزمون فرضیه‌های پژوهش و مقایسه عملکرد روش‌های مختلف پیش‌بینی و همچنین مقایسه سودمندی روش‌های مختلف انتخاب متغیرهای پیش‌بین، از نرم‌افزار SPSS نسخه ۲۶ استفاده شده است.

### روش‌های انتخاب ویژگی: تحلیل مولفه‌های اصلی و ریلیف

**روش‌های یادگیری ماشین:** رگرسیون لجستیک، درخت تصمیم C 5.0، رگرسیون موزون محلی، روش تجمیعی (روش آدابوست)، ماشین بردارهای پشتیبان (روش هسته تابع پایه شعاع)، روش ترکیبی آدابوست ماشین بردار پشتیبان

### روش تجزیه و تحلیل داده‌ها و آزمون فرضیه‌ها

پس از مشخص شدن متغیرهای پیش‌بین به وسیله هر یک از روش‌های انتخاب متغیر، از روش‌های یادگیری ماشین فوق الذکر برای پیش‌بینی مدیریت سود استفاده شده است. به منظور ارزیابی عملکرد هر یک از روش‌های یادگیری ماشین برای پیش‌بینی مدیریت سود از معیارهای ارزیابی میانگین دقت، خطاهای نوع اول و خطاهای نوع دوم استفاده می‌شود.

در پژوهش حاضر برای طبقه بندی درجه مدیریت سود اقلام تعهدی از گره گسسته سازی استفاده شده است. در این گره، از تکنیک میانگین و انحراف معیار برای گسسته سازی استفاده می‌شود که به این صورت عمل می‌کند از مجموعه نمونه‌ها که در اینجا اقلام تعهدی اختیاری است، میانگین و انحراف معیار آن‌ها را محاسبه می‌کند و سپس در مرحله اول یک درجه از مقدار انحراف معیار را به مقدار میانگین اضافه می‌کند، مقدار میانگین به علاوه یک درجه انحراف معیار را به عنوان سقف (مدیریت سود بسیار رو به بالا) در نظر می‌گیریم؛ و در مرحله دوم؛ یک درجه از مقدار انحراف معیار را از مقدار میانگین کم می‌کند، مقدار میانگین منهای یک درجه انحراف معیار را به عنوان کف (مدیریت سود بسیار رو به پایین) در نظر می‌گیریم. با این حال مقداری از اقلام تعهدی اختیاری که بین سقف و کف باشد، رفتار مدیریت سود جزئی (اندک) تلقی می‌شود. برای ارزیابی عملکرد هر یک از روش‌های مورد استفاده برای تشخیص مدیریت سود، از ماتریس دقت پیش‌بینی نشان داده شده در جدول ۳-۲ استفاده می‌شود. در این ماتریس، خطای نوع اول به وضعیتی اشاره دارد که مدیریت سود واقعاً "بسیار رو به پایین" یا وضعیت "بسیار رو به بالا" است، اما طبقه بندی کننده وضعیت خروجی نادرست مدیریت سود جزئی را پیش‌بینی می‌کند. در مقابل، یک خطای نوع ۲ نشان می‌دهد که وضعیت مدیریت سود نسبتاً جزئی به اشتباه به عنوان بسیار رو به بالا یا پایین طبقه بندی می‌شود. رابطه های ۳-۳۲ تا ۳-۳۴ به

ترتیب فرمول های محاسبه خطای نوع اول، خطای نوع ۲ و دقت متوسط را نشان می دهد. با توجه به طبقه بندی درجه مدیریت سود ارقام تعهدی، خطای نوع اول بسیار مهمتر از خطای نوع دوم است زیرا ما علاقه مندیم فعالیت های مدیریت سود بسیار رو به بالا/پایین را ضبط و شناسایی کنیم زیرا ممکن است برای شرکت ها و سهامداران پر ریسک تر از مدیریت سود متوسط باشد.

رابطه (۳-۳۲)

$$TypeI = \frac{FMUD}{FMUD + TEUD}$$

رابطه (۳-۳۳)

$$TypeII = \frac{FEUD}{FEUD + TMUD}$$

رابطه (۳-۳۴)

$$Average\ accuracy = \frac{TEUD + TMUD}{FEUD + TMUD + TEUD + FMUD}$$

جدول ۳-۲: ماتریس دقت پیش بینی

وضعیت واقعی		وضعیت مدیریت سود
وضعیت نسبتاً جزئی	وضعیت بسیار رو به بالا یا بسیار رو به پایین	
وضعیت نادرست بسیار رو به بالا یا بسیار رو به پایین (FEUD <sup>۲</sup> )	وضعیت درست بسیار رو به بالا یا بسیار رو به پایین (TEUD <sup>۱</sup> )	وضعیت بسیار رو به بالا یا بسیار رو به پایین
وضعیت درست نسبتاً جزئی (TMUD <sup>۴</sup> )	وضعیت نادرست نسبتاً جزئی (FMUD <sup>۳</sup> )	وضعیت نسبتاً جزئی

همچنین به منظور ارزیابی بیشتر نیز از آزمون t زوجی (بر اساس صد دقت حاصل از اجرای روایی متقابل ۱۰ بخشی با ۱۰ بار تکرار در هر روش پیش بینی) استفاده شده است.

در پژوهش حاضر به منظور بررسی تعمیم پذیری پیش بینی های انجام شده از روایی متقابل ۱۰ بخشی استفاده می شود. روایی متقابل ۱۰ بخشی برای برآورد نرخ خطای واقعی کاملاً قابل اتکا و کافی است

1. True extremely upward or downward

2. False extremely upward or downward

3. False moderately upward or downward

4. True moderately upward or downward

# روش اجرای پژوهش

