

ỨNG DỤNG CÔNG CỤ WEKA ĐỂ ĐÁNH GIÁ HIỆU QUẢ HOẠT ĐỘNG CỦA DOANH NGHIỆP KINH DOANH BẤT ĐỘNG SẢN VIỆT NAM

Nguyễn Thị Tuyết Như^{1,*}, Trần San Đào¹, Lý Phát Cường¹, Phan Minh Đạt¹,
Phan Tấn Anh Phong², Đinh Hùng³, Phạm Minh Dũng³

¹ Khoa Kinh tế - Tài chính, Trường Đại học Ngoại ngữ - Tin học TP.HCM

² Khoa Ngôn ngữ và văn hóa Phương Đông, Trường ĐH Ngoại ngữ - Tin học TP.HCM

³ Khoa Công nghệ thông tin, Trường Đại học Ngoại ngữ - Tin học TP.HCM

{*nhu.ntt, dao.ts, cuong.lp*}@huflit.edu.vn, *dat.m.phan@gmail.com*, {*phongphan, dinhhung621, pmdung*}@huflit.edu.vn

TÓM TẮT— Bài này trình bày ứng dụng weka để lựa chọn tiêu chí đánh giá hiệu quả hoạt động của doanh nghiệp kinh doanh bất động sản Việt Nam giai đoạn 2015-2019. Từ kết quả ứng dụng WEKA, nhóm tác giả phân tích và luận bàn về độ tin cậy của kết quả ứng dụng, rút ra những nhận xét ưu và nhược điểm để làm bài học kinh nghiệm cho những ứng dụng WEKA tiếp sau.

Từ khóa— Fintech, WEKA, kinh doanh bất động sản.

I. ĐẶT VẤN ĐỀ

“Bất động sản là kênh trú ẩn an toàn cho tiền nhàn rỗi” là tựa đề bài viết đăng trên báo Công an nhân dân ngày 31/07/2020 [1]. Qua đó, các chuyên gia kinh tế đã nhận định trong bối cảnh biến động của giá vàng, tỷ giá, lãi suất, chứng khoán... thì bất động sản đang là kênh đầu tư an toàn và tiếp tục “hái” ra tiền. Tại Việt Nam hiện nay, đầu tư bất động sản được xem là kênh đầu tư “trú ẩn” lạm phát hữu hiệu, đã thu hút đầu tư từ khoản tiền nhàn rỗi của các tổ chức, cá nhân đến từ nhiều ngành nghề khác nhau. Trước khi ra quyết định đầu tư, nhà đầu tư phải căn cứ trên các thông tin tài chính doanh nghiệp kinh doanh bất động sản để đánh giá hiệu quả hoạt động.

Tuy nhiên, với phương pháp thống kê thủ công, nhà đầu tư khó thực hiện trên tất cả thông tin tài chính của các doanh nghiệp và phải mất nhiều thời gian tham chiếu tính toán với độ chính xác không cao. Ứng dụng công nghệ vào lĩnh vực tài chính, nhà đầu tư có thể khảo sát thông tin trên vùng dữ liệu lớn một cách nhanh chóng, chính xác để kịp thời ra quyết định đầu tư, nắm bắt cơ hội đầu tư.

Giai đoạn (2015 – 2019) là giai đoạn mà thị trường bất động sản Việt Nam phát triển tương đối ổn định. Ứng dụng công cụ WEKA để lựa chọn tiêu chí đánh giá hiệu quả hoạt động của doanh nghiệp kinh doanh bất động sản Việt Nam trong giai đoạn (2015 – 2019) để tư vấn đầu tư, lựa chọn và ra quyết định đầu tư, đây cũng là một trong những ứng dụng công nghệ tài chính phù hợp trong thời đại công nghiệp 4.0.

II. CÁC KHÁI NIỆM

Hiệu quả hoạt động của doanh nghiệp được minh chứng qua nhiều loại chỉ số có liên quan đến nguồn lực của doanh nghiệp. Trong phạm vi nghiên cứu này, nhóm tác giả chọn các chỉ tiêu cơ bản như sau: lợi nhuận sau thuế/doanh thu; lợi nhuận trên tài sản – ROA; lợi nhuận trên vốn chủ sở hữu – ROE; tăng trưởng doanh thu hàng năm; thu nhập trên mỗi cổ phiếu – EPS; lợi nhuận trên vốn đầu tư – ROIC.

A. LỢI NHUẬN SAU THUẾ/ DOANH THU

Lợi nhuận sau thuế/ doanh thu, còn được gọi là tỷ suất lợi nhuận trên doanh thu, là một tỷ số tài chính dùng để theo dõi tình hình sinh lợi của doanh nghiệp, chỉ tiêu này phản ánh quan hệ giữa lợi nhuận ròng và doanh thu của doanh nghiệp [2].

Tỷ suất lợi nhuận trên doanh thu trong một kỳ nhất định được tính bằng cách lấy lợi nhuận ròng hoặc lợi nhuận sau thuế trong kỳ chia cho doanh thu trong kỳ. Đơn vị tính là %. Cả lợi nhuận ròng lẫn doanh thu đều có thể lấy từ báo cáo kết quả kinh doanh của công ty.

Công thức tính tỷ số này như sau:

$$\text{Tỷ số lợi nhuận trên doanh thu} = 100 \times \frac{\text{Lợi nhuận ròng (hoặc lợi nhuận sau thuế)}}{\text{Doanh thu}}$$

Tỷ suất này cho biết lợi nhuận chiếm bao nhiêu phần trăm trong doanh thu. Tỷ số này mang giá trị dương nghĩa là doanh nghiệp kinh doanh có lãi; tỷ số càng lớn nghĩa là lãi càng lớn. Tỷ số mang giá trị âm nghĩa là doanh nghiệp kinh doanh thua lỗ.

* Tác giả điều phối (Corresponding Author)

Tuy nhiên, tỷ số này phụ thuộc vào đặc điểm kinh doanh của từng ngành. Vì thế, khi theo dõi tình hình sinh lợi của doanh nghiệp, người ta so sánh tỷ số này của doanh nghiệp với tỷ số bình quân của toàn ngành mà doanh nghiệp đó tham gia. Mặt khác, tỷ số này và số vòng quay tài sản có xu hướng ngược nhau. Do đó, khi đánh giá tỷ số này, người phân tích tài chính thường tìm hiểu nó trong sự kết hợp với số vòng quay tài sản.

B. TỶ LỆ TĂNG TRƯỞNG DOANH THU

Tỷ lệ tăng trưởng doanh thu còn có tên tiếng Anh là Sale Growth Ratio, "Revenue Growth Rate", "Revenue Growth Ratio". Tỷ lệ tăng trưởng doanh thu cho biết mức tăng trưởng doanh thu tương đối (tính theo phần trăm) qua các năm. Tỷ lệ này nhỏ hơn không đồng nghĩa với tăng trưởng âm [3].

$$\text{Tính tỷ lệ tăng trưởng doanh thu hàng năm} = \left[\left(\frac{f}{s} \right)^{1/y} - 1 \right] \times 100$$

Trong đó:

- f: Doanh thu năm cuối kỳ;
- s: Doanh thu năm đầu kỳ;
- y: số năm trong kỳ.

Doanh nghiệp có tỷ lệ tăng trưởng doanh thu cao thường đang trong giai đoạn phát triển mạnh, thị phần tăng hoặc đang mở rộng kinh doanh sang các thị trường, hoặc lĩnh vực mới. Tuy nhiên tỷ lệ tăng trưởng doanh thu cao không nhất thiết đi kèm với tỷ lệ tăng trưởng lợi nhuận cao. Tùy vào xu hướng của tỷ lệ tăng trưởng doanh thu mà mức tăng trưởng được đánh giá là bền vững, không ổn định, phi mã hay tuột dốc.

Những doanh nghiệp có mức tăng trưởng doanh thu ổn định ở mức cao luôn được các nhà đầu tư đặc biệt quan tâm.

C. LỢI NHUẬN TRÊN TÀI SẢN - ROA

Chỉ số ROA (Return on Asset) là chỉ số tài chính cho biết tỷ suất lợi nhuận trên tài sản, thể hiện mối tương quan giữa mức sinh lợi nhuận của công ty so với tài sản mà công ty đó đã có và đầu tư. Chỉ số ROA này cho thấy hiệu quả công ty trong việc đầu tư qua những tài sản và thu lại lợi nhuận như thế nào [2].

Công thức tính chỉ số ROA:

$$\text{ROA} = \frac{\text{Lợi nhuận sau thuế}}{\text{Tổng tài sản bình quân}} \times 100\%$$

Chỉ số ROA cho ta biết được doanh nghiệp kiếm được bao nhiêu tiền và hưởng lợi nhuận là bao nhiêu trên 1 đồng tài sản. Bởi vì tài sản của một công ty được hình thành từ vốn vay và vốn chủ sở hữu, do đó ROA chính là thước đo hiệu quả của việc chuyển hóa số vốn đầu tư thành lợi nhuận.

Chỉ số ROA càng cao cho thấy công ty đang kiếm được nhiều tiền hơn trên số tiền đầu tư ít hơn. Do đó trên sàn chứng khoán, cổ phiếu của doanh nghiệp nào có chỉ số ROA cao luôn được ưa chuộng hơn.

D. LỢI NHUẬN TRÊN VỐN CHỦ SỞ HỮU - ROE

Chỉ số ROE (Return On Equity) là chỉ số tài chính cho biết tỷ suất lợi nhuận ròng trên vốn chủ sở hữu. Nên đôi khi chỉ số này còn được gọi nhanh là chỉ số lợi nhuận trên vốn chủ. Đây là một trong những chỉ số quan trọng, đo lường mức độ hiệu quả trong việc sử dụng vốn chủ sở hữu của doanh nghiệp [2].

Công thức tính chỉ số ROE:

$$\text{ROE} = \frac{\text{Lợi nhuận sau thuế kỳ này}}{\text{Vốn chủ sở hữu bình quân}} \times 100\%$$

Nhà đầu tư đặc biệt chú ý tới chỉ số ROE, vì chỉ số này cho biết với số vốn mà chủ sở hữu bỏ ra thì doanh nghiệp sẽ thu được bao nhiêu phần trăm lợi nhuận. Doanh nghiệp có ROE cao và luôn duy trì được ở mức ổn định, đồng nghĩa với việc doanh nghiệp đã sử dụng vốn một cách hiệu quả.

E. THU NHẬP TRÊN MỖI CỔ PHIẾU - EPS

Chỉ số EPS (Earning Per Share) là chỉ số tài chính cho biết thu nhập trên mỗi cổ phiếu. Có nhiều cách hiểu khác nhau về EPS, nó có thể là một biến số quan trọng khi tính toán giá cổ phiếu, hay là một chỉ số tài chính quan trọng khi đầu tư chứng khoán, hoặc có thể là lãi cơ bản trên mỗi cổ phiếu. Nhìn chung, nhà đầu tư có thể hiểu chỉ số này là phần lợi nhuận sau thuế trên mỗi cổ phiếu thường của các cổ đông, sau khi đã trừ đi các khoản cổ tức ưu đãi. Chỉ số này đóng vai trò quan trọng trong việc cấu thành nên hệ số P/E. [2].

Để tính được chỉ số EPS cơ bản của một công ty, doanh nghiệp thì cần phải có bảng cân đối kế toán và báo cáo kết quả kinh doanh. Những số liệu để thiết lập công thức tính bao gồm:

- Tổng số cổ phiếu thường đang lưu hành bình quân
- Lợi nhuận sau thuế

- Chi trả cổ tức ưu đãi (nếu có)

Công thức tính chỉ số EPS:

$$EPS = \frac{\text{Lợi nhuận ròng sau thuế} - \text{Cổ tức cổ phiếu ưu đãi}}{\text{Tổng số cổ phiếu thường đang lưu hành bình quân}}$$

EPS là một chỉ số tổng hợp phản ánh kết quả kinh doanh của doanh nghiệp. Từ đó, giúp các nhà đầu tư có thể dễ dàng hiểu và so sánh giữa các loại cổ phiếu. Thường thì những doanh nghiệp nào có EPS cao hơn sẽ thu hút được nhiều sự đầu tư hơn.

F. F. LỢI NHUẬN TRÊN VỐN ĐẦU TƯ - ROIC

ROIC (Return on Invested Capital) là một chỉ số tài chính cho biết tỷ suất lợi nhuận trên vốn đầu tư, thể hiện mức độ sinh lời trên một đồng vốn đầu tư. Chỉ số này giúp các cổ đông thấy rõ được giá trị của doanh nghiệp tăng hay giảm.

Công thức tính chỉ số ROIC:

$$ROIC = \frac{\text{Lợi nhuận sau thuế kỳ này}}{\text{Tổng vốn chủ sở hữu + nợ dài hạn}} \times 100\%$$

Doanh nghiệp được đánh giá là đang hoạt động có tăng trưởng khi ROIC $\geq 10\%$. Nếu thấy chỉ số ROIC tăng liên tục theo từng năm và luôn ở trên mức 10% thì có thể đầu tư.

III. PHƯƠNG PHÁP LỰA CHỌN TIÊU CHÍ ĐÁNH GIÁ HIỆU QUẢ HOẠT ĐỘNG

Để lựa chọn tiêu chí đánh giá hiệu quả hoạt động của doanh nghiệp đầu tư bất động sản Việt Nam trong giai đoạn (2015 – 2019), nhóm tác giả phân công thực hiện các phần hành như sau:

- Tải thông tin từ <https://www.cophieu68.vn> về báo cáo tài chính của 64 doanh nghiệp Việt Nam trong khối ngành đầu tư bất động sản trong 5 năm (2015-2019); sử dụng bảng số liệu chính thức bao gồm 320 bộ chỉ số về các chỉ tiêu tài chính: ROA, ROE, tăng trưởng doanh thu hàng năm, tăng trưởng EPS hàng năm.
- Định dạng lại dữ liệu sang CSV (Phụ lục 1. Định dạng 295 bộ chỉ số huấn luyện; phụ lục 2. Định dạng 25 bộ chỉ số tập thử)
- Tập hợp và dùng Pycharm rút ra các chỉ số tài chính mong muốn, cụ thể là từ dòng 13 đến 18 tương ứng trong từng file.
- Tiến hành phân định các bộ chỉ tiêu vào 3 nhóm good - fair - bad (hoạt động tốt - trung bình - kém) như sau:
 - Dùng chức năng cluster trong Weka để rút ra gợi ý về phân bố lớn đến nhỏ trong các chỉ tiêu. Tại bước này nhóm chia ngẫu nhiên bộ số liệu thành 24 nhóm. Từ gợi ý phân loại và giá trị trung bình trong từng nhóm của weka, nhóm có thể có cái nhìn tổng quan về top những bộ chỉ số tốt nhất, kém nhất và trung bình của ngành.
 - Sau đó, nhóm tiến hành xem xét từng bộ chỉ tiêu mỗi năm của từng doanh nghiệp để tiến hành gán lần lượt các mức phân loại good - fair - bad
 - Vì mẫu khá nhỏ, nhóm chọn 295 bộ số liệu làm tập huấn luyện trong ứng dụng Classify (tree - J48) trong weka. Còn lại 25 bộ số liệu sẽ được sử dụng làm tập thử.
- Thảo luận nghiên cứu
- Nhận xét, kết luận

IV. KẾT QUẢ THỬ NGHIỆM VÀ BÀN LUẬN

Khi chạy cluster trên bộ số liệu [3], chia thành 24 nhóm và ra được kết quả theo bảng bên dưới. Căn cứ theo giá trị bình quân của các nhóm do Weka gán, có thể phân định được 3 mức hiệu quả hoạt động tốt – trung bình – kém phù hợp hơn với tình hình chung của ngành. Kết quả thể hiện qua Bảng 1.

Sau đó, quay lại bảng số liệu chính thức bao gồm 320 bộ chỉ số, tiến hành gán các giá trị (good, fair, bad) cho từng bộ chỉ số. Tiếp theo, chia bộ chỉ số thành 2 tập tin: tập huấn luyện (295 bộ chỉ số) và tập thử (25 bộ chỉ số). Sau khi đưa Tập huấn luyện vào Weka, dùng Classiy (Tree => K48) trả về kết quả như Bảng 2.

Bảng 1. Bảng biểu hướng dẫn

Attribute	Full	0	1	2	3	4	5	6	7	8	9	10	11
	-320	-1	-1	-5	-7	-46	-5	-4	-1	-6	-1	-31	-1
pre tax	0	-0.2	-50	0.74	-0.7	0.03	0.49	-1	3.47	0.41	-0.3	0.17	24.4
after tax	-0.1	-0.2	-50	0.62	-0.7	0.01	0.43	-1	3.47	0.37	-0.3	0.14	24.4
roa	0.03	-0.1	0	0.03	-0.2	0.01	0.21	-0.1	0.61	0.03	-0.3	0.03	0.01
roe	0.06	2.08	-0.1	0.08	-0.8	0.01	0.48	-0.1	0.79	0.08	-0.8	0.11	0.02
roic	0.1	-0.3	-60	1.02	-0.7	0.01	1.18	-1	19.6	0.46	-0.2	0.2	11.8
growth rat	7.04	-0.1	-0.9	-0.4	-0.1	0.37	2.8	-0.1	40.8	0.14	2036	0.28	-0.4
esp	0.12	0	-3.2	-0.1	-1.1	-0.5	1.22	-22	0	5.38	-26	0.19	0
note		bad	bad	bad	bad	bad	good	bad	good	fair	bad	fair	bad
Attribute	Full	12	13	14	15	16	17	18	19	20	21	22	23
	-320	-9	-2	-2	-2	-3	-13	-52	-24	-12	-8	-29	-55
pre tax	0	0.35	-13	2.08	-1.4	-7.5	0.19	0.3	0.34	0.1	-0.2	0.45	0.12
after tax	-0.1	0.28	-13	1.88	-1.9	-7.1	0.17	0.3	0.32	0.08	-0.2	0.39	0.1
roa	0.03	0.08	-0.4	0.02	-0.5	0	0.04	0.1	0.13	0.02	-0.1	0.09	0.02
roe	0.06	0.21	-4.5	0.06	-0.7	0.27	0.11	0.1	0.23	0.05	0.08	0.21	0.06
roic	0.1	0.42	-3.9	2.94	-2.9	-5.4	0.26	0.4	0.64	0.11	-0.2	0.8	0.14
growth rat	7.04	2.59	-0.6	-0.8	-0.2	-0.6	0.4	0.4	0.26	0.78	0.4	0.66	1.08
esp	0.12	18.9	1.7	0.97	-57	-1.4	1.35	0.1	0.5	2.48	-0.4	0.6	0.09
note		good	bad	fair	bad	bad	fair	good	good	fair	bad	good	fair

Bảng 2. Kết quả đưa tập huấn luyện vào Weka, dùng Classiy (Tree => K48)

```
=== Run information ===
```

```
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2 Relation: trail3_train295_6facts_2
```

```
Instances: 295
```

```
Attributes: 7
```

```
after tax roa
```

```
roe roic growth esp class
```

```
Test mode: evaluate on training data
```

```
=== Classifier model (full training set) ===
```

```
J48 pruned tree
```

```
roa <= 0.02
```

```
| roe <= 0.0439: bad (90.0)
```

```
| roe > 0.0439
```

```
|| roa <= 0.01: bad (11.0/1.0)
```

```
|| roa > 0.01
```

```
||| roic <= 0.12
```

```
|||| roe <= 0.071: bad (9.0)
```

```
|||| roe > 0.071: fair (7.0/1.0)
```

```

||| roic > 0.12: fair (16.0/1.0) roa > 0.02
|
|         roe <= 0.1832
|
|         | roe <= 0.1534: fair (92.0)
|         | roe > 0.1534
|         || ->after tax <= 0.44
|         ||| roa <= 0.13: fair (19.0/1.0)
|         ||| roa > 0.13: good (2.0)
|         || ->after tax > 0.44: good (3.0)
|         roe > 0.1832
|         | esp <= -0.22: fair (4.0/1.0)
|         | esp > -0.22
|         || roa <= 0.06
|         ||| growth <= 0.36: good (2.0)
|         ||| growth > 0.36: fair (2.0)
|         || roa > 0.06: good (38.0)
Number of Leaves : 13 Size of the tree : 25
Time taken to build model: 0.02 seconds
=== Evaluation on training set ===
Time taken to test model on training data: 0.01 seconds
=== Summary ===
Correctly Classified Instances 290 98.3051 %
Incorrectly Classified Instances 5 1.6949 %

F-Measure MCC ROC Area PRC Area Class
0.957 0.000 1.000 0.957 0.978 0.975 0.999 0.992 good
0.982 0.005 0.991 0.982 0.986 0.978 0.999 0.997 bad
0.993 0.025 0.971 0.993 0.982 0.966 0.997 0.995 fair
Weighted Avg. 0.983 0.014 0.983 0.983 0.983 0.972 0.998 0.995
=== Confusion Matrix ===
a b c <-- classified as
45 0 2 | a = good 0 109 2 | b = bad 0 1 136 | c = fair
    
```

Ứng dụng phần mềm WEKA trên tập thử (25 bộ chỉ số), dùng Classiy (Tree => K48). (5 good - 10 fair - 10 bad).
Danh sách các bộ chỉ số được lựa chọn theo Bảng 3.

Bảng 3. Danh sách các bộ chỉ số được lựa chọn chạy tập thử

company	after tax	roa	roe	roic	growth	Esp	class
CEO	0.25	0.08	0.2076	0.42	0.14	1.01	good
DXG	0.4	0.16	0.3131	0.75	1.67	1.01	good
IDV	0.73	0.13	0.4682	2.88	1.32	0.83	good
IDV	0.67	0.15	0.6127	2.42	1.77	1.67	good
D2D	0.49	0.18	0.4707	1.3	1.7	3.16	good

FLC	0.17	0.06	0.1237	0.23	0.15	0.15	fair
DRH	0.65	0.08	0.1109	0.77	-0.42	3.51	fair
LGL	0.06	0.04	0.0992	0.08	0.1	-0.36	fair
KBC	0.41	0.04	0.0737	0.69	0.34	0.93	fair
IJC	0.19	0.03	0.1342	0.3	0.33	0.24	fair
IDI	0.04	0.05	0.1172	0.05	0.22	-0.49	fair
NTL	0.14	0.06	0.1018	0.19	1.37	0.11	fair
NLG	0.17	0.07	0.1366	0.25	1	0.67	fair
REE	0.35	0.1	0.1381	0.53	0.01	-0.2	fair
TDC	0.08	0.02	0.1069	0.11	0.28	0.01	fair
HAG	-1.85	-0.1	-0.2281	-2.08	-0.61	-66.2	bad
DTA	0.01	0.0006	0.0013	0.01	-0.22	0.77	bad
DLR	0.01	0.01	0.0246	0.01	1.3	0	bad
ITA	0.01	0.0019	0.0008	0.02	1.48	-0.02	bad
IDJ	0.0024	0.0001	0.0001	0	-0.49	0	bad
PFL	-0.2	-0.04	-0.0578	-0.21	6.51	0	bad
SHN	0.02	0.02	0.0418	0.02	1.84	1.19	bad
SJS	-7.65	0.02	0.0678	-0.75	-0.91	-1	bad
UDC	0.01	0.01	0.0716	0.01	-0.11	-0.56	bad
CLG	0	0.0008	0.0032	0.008	0.11	-0.46	bad

Ứng dụng phần mềm WEKA trên tập thử (25 bộ chỉ số), dùng Classiy (Tree => K48), kết quả theo bảng 4.

Bảng 4. Kết quả đưa tập thử vào Weka, dùng Classiy (Tree => K48)

```

=== Run information ===
Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W weka.classifiers.trees.J48 -- - C 0.25 -M 2
Relation: trail3_train295_6facts_2 Instances: 295
Attributes: 7
after tax roa
roe roic growth esp class
Test mode: user supplied test set: size unknown (reading incrementally)
=== Classifier model (full training set) === InputMappedClassifier:
J48 pruned tree

roa <= 0.02
| roe <= 0.0439: bad (90.0)
| roe > 0.0439
|| roa <= 0.01: bad (11.0/1.0)
|| roa > 0.01
||| roic <= 0.12
||| roe <= 0.071: bad (9.0)
||| roe > 0.071: fair (7.0/1.0)
||| roic > 0.12: fair (16.0/1.0) roa > 0.02
| roe <= 0.1832
|| roe <= 0.1534: fair (92.0)
|| roe > 0.1534
||| ->after tax <= 0.44
||| roa <= 0.13: fair (19.0/1.0)
||| roa > 0.13: good (2.0)
||| ->after tax > 0.44: good (3.0)
| roe > 0.1832
|| esp <= -0.22: fair (4.0/1.0)
|| esp > -0.22
||| roa <= 0.06
||| growth <= 0.36: good (2.0)
||| growth > 0.36: fair (2.0)
||| roa > 0.06: good (38.0) Number of Leaves : 13
Size of the tree : 25

```

Attribute mappings:

Model attributes Incoming attributes

(numeric) after tax --> 1 (numeric) after tax (numeric) roa --> 2 (numeric) roa
(numeric) roe --> 3 (numeric) roe (numeric) roic --> 4 (numeric) roic
(numeric) growth --> 5 (numeric) growth (numeric) esp --> 6 (numeric) esp
(nominal) class --> 7 (nominal) class Time taken to build model: 0.01 seconds

=== Evaluation on test set ===

Time taken to test model on supplied test set: 0 seconds

=== Summary ===

Correctly Classified Instances 25 100 %
Incorrectly Classified Instances 0 0 %

Kappa statistic 1

Mean absolute error 0.0062

Root mean squared error 0.0277

Relative absolute error 1.479 %

Root relative squared error 5.9585 %

Total Number of Instances 25

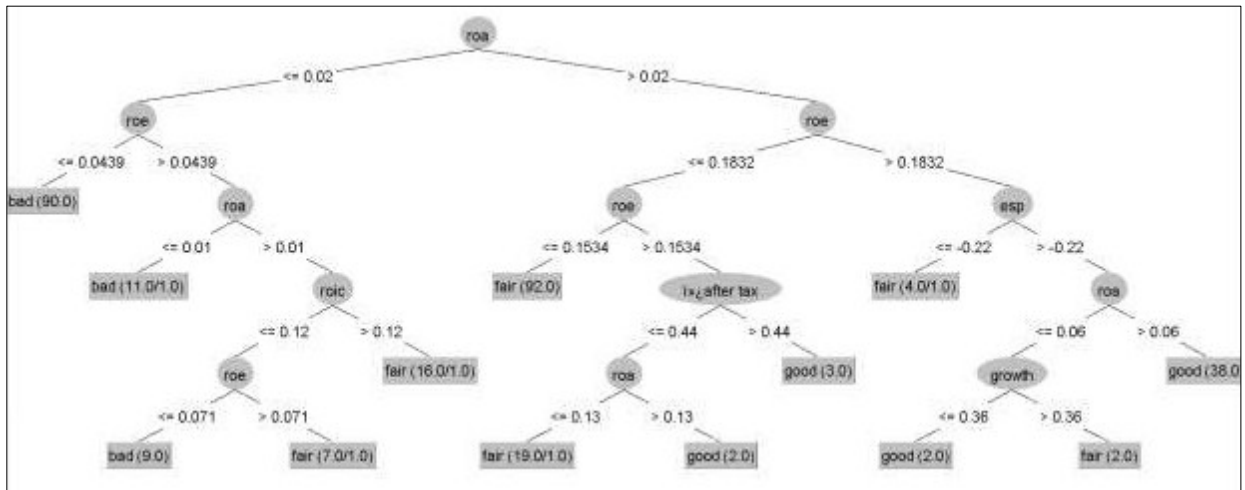
=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC
1.000	0.000	1.000	1.000	1.000	1.000	1.000	good
1.000	0.000	1.000	1.000	1.000	1.000	1.000	bad
1.000	0.000	1.000	1.000	1.000	1.000	1.000	fair
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000

=== Confusion Matrix === a b c <-- classified as

5 0 0 | a = good 0 10 0 | b = bad 0 0 10 | c = fair

Kết quả ước lượng được biểu diễn qua cây quy luật trong Hình 1.



Hình 1. Cây quy luật về tiêu chí đánh giá hiệu quả hoạt động của các doanh nghiệp bất động sản từ 2015- 2019

A. PHÂN TÍCH KẾT QUẢ THỬ NGHIỆM

Tiêu chí đánh giá hiệu quả hoạt động của các doanh nghiệp Việt Nam kinh doanh bất động sản từ 2015- 2019 qua cây quy luật được hiểu như sau:

Phía bên trái:

- Doanh nghiệp có chỉ số: $ROA \leq 0,02$ và $ROE \leq 0,0439$ thì kém
- Doanh nghiệp có chỉ số: $0,01 < ROA \leq 0,02$; và $ROE > 0,0439$ và $ROIC > 0,12$: trung bình
- Doanh nghiệp có chỉ số: $0,01 < ROA \leq 0,02$; $ROIC \leq 0,12$ và $ROE > 0,071$: trung bình
- Doanh nghiệp có chỉ số: $0,01 < ROA \leq 0,02$; $ROIC \leq 0,12$ và $ROE < 0,071$: kém

Phía bên phải

- Doanh nghiệp có chỉ số: $ROA > 0,02$; $0,01534 \leq ROE < 0,1832$: trung bình
- Doanh nghiệp có chỉ số: $ROA > 0,02$; $0,01534 < ROE < 0,1832$ và lợi nhuận sau thuế $> 0,44$ thì tốt

- Doanh nghiệp có chỉ số: $ROA > 0,13$; $0,01534 < ROE < 0,1832$; lợi nhuận sau thuế $\leq 0,44$ thì tốt
- Doanh nghiệp có chỉ số: $ROA \leq 0,13$; $0,01534 < ROE < 0,1832$; lợi nhuận sau thuế $\leq 0,44$ thì trung bình
- Doanh nghiệp có chỉ số: $ROA > 0,06$; $ROE > 0,1832$ và $esp > -0,22$ thì tốt
- Doanh nghiệp có chỉ số: $ROA \leq 0,06$; $ROE > 0,1832$, $esp > -0,22$ và $growth > 0,36$ thì trung bình
- Doanh nghiệp có chỉ số: $ROA \leq 0,06$; $ROE > 0,1832$, $esp > -0,22$ và $growth \leq 0,36$ thì good

B. THẢO LUẬN KẾT QUẢ THỬ NGHIỆM

Ứng dụng phần mềm WEKA trên Cluster để lựa chọn thuộc tính trên vùng dữ liệu xác định là một trong những ứng dụng phân tích tài chính doanh nghiệp hiện đại phù hợp với công nghệ ngày nay. Tuy nhiên, trong bất kỳ mô hình ước lượng nào cũng đều có những ưu điểm và hạn chế nhất định.

Ưu điểm:

- Ứng dụng phần mềm WEKA để lựa chọn thuộc tính trên vùng dữ liệu lớn một cách nhanh chóng, chính xác, điều mà không dễ dàng thực hiện được với phương pháp thủ công truyền thống.
- Ứng dụng phần mềm WEKA để lựa chọn thuộc tính trên vùng dữ liệu lớn một cách khách quan, khoa học, không lệ thuộc vào cảm tính.
- Ứng dụng phần mềm WEKA khi cho kết quả nghiên cứu có đánh giá mức độ chính xác của mô hình (trong nghiên cứu này là 98.3051 %, sai số 1.6949 %)

Hạn chế:

- Theo như phân tích cây quy luật ở trên, mô hình vẫn chứa một điểm khá vô lý tại nhánh trái khi Growth thấp lại được phân loại thành Good trong khi cao lại phân thành Fair. Ngoài ra, mô hình vẫn chưa được xem là hoàn thiện và cần phát triển thêm. Nguyên nhân có thể xuất phát những lý do sau đây:
- Mẫu huấn luyện còn khá nhỏ, chỉ 295 bộ số liệu.
- Số lượng các bộ chỉ tiêu tài chính được phân loại vào 3 nhóm cũng chưa lý tưởng. Cụ thể 47-137-111 bộ chỉ số ứng với từng nhóm Good - Fair - Bad. Mô hình huấn luyện lý tưởng nên có số mẫu bằng nhau ở từng nhóm. Đây một phần xuất phát từ lý do khách quan thị trường, tình hình hoạt động kinh doanh của các doanh nghiệp trong mẫu.
- Việc phân loại Good - Fair - Bad vẫn còn nhiều yếu tố chủ quan. Đối với những trường hợp số liệu rất cao hoặc rất thấp, việc phân loại là vô cùng dễ dàng, tuy nhiên, có khá nhiều trường hợp doanh nghiệp đạt kết quả rất tốt ở một số chỉ tiêu, nhưng lại khá thấp ở những chỉ tiêu còn lại. Do đó, việc đánh giá trở nên chủ quan và gây ảnh hưởng lên quá trình rút ra kết luận của phần mềm.
- Số lượng các chỉ số tài chính có thể đưa vào đánh giá tình hình kinh doanh của một doanh nghiệp là khá nhiều, do đó, việc lựa chọn các chỉ số được dùng trong mô hình này cũng có sự cân nhắc chủ quan.
- Chỉ số Growth (tăng trưởng doanh thu) biến động khá lớn qua các năm ở các doanh nghiệp, điều này có thể gây ảnh hưởng đến kết luận của phần mềm. Có khá nhiều doanh nghiệp dù rằng tốc độ tăng trưởng không cao nhưng những chỉ số khác đều rất tốt nên theo nguyên tắc vẫn được xếp vào nhóm Good, trong khi có một số doanh nghiệp có tăng trưởng doanh thu cao, nhưng chỉ là phản ánh sự phục hồi từ tình trạng hoạt động kém hiệu quả trước đó, nên chỉ được xếp vào nhóm Fair.

Tất cả những hạn chế này có thể được khắc phục, nếu khảo sát thêm nhiều bộ chỉ số nữa.

V. KẾT LUẬN

Ứng dụng phần mềm WEKA trên Cluster để lựa chọn thuộc tính trên vùng dữ liệu lớn là một trong những ứng dụng công nghệ để phân tích tài chính doanh nghiệp phù hợp với công nghệ ngày nay. Nhóm tác giả đã vận dụng WEKA để xác định tiêu chí đánh giá hiệu quả hoạt động của doanh nghiệp kinh doanh bất động sản Việt Nam trong giai đoạn (2015 – 2019). Đây cũng là một trong những ứng dụng của Fintech trong lĩnh vực tài chính và đầu tư.

VI. TÀI LIỆU THAM KHẢO

- [1] An An (2020), Bất động sản là kênh trú ẩn an toàn cho tiền nhàn rỗi, <http://cand.com.vn/Kinh-te/Bat-dong-san-la-kenh-tru-an-an-toan-cho-tien-nhan-roi-605229>, Accessed date: Mar 3, 2021
- [2] Anonymous, Tài chính doanh nghiệp, https://vi.wikipedia.org/wiki/Th%E1%BB%83_lo%E1%BA%A1:T%C3%A0i_ch%C3%ADnh_doanh_nghi%E1%BB%87p, Accessed date: Mar 3, 2021
- [3] Đỗ Phúc, Phan Hồng Trung (2021), Bài giảng khóa Bồi dưỡng phương pháp TTNT ứng dụng trong quản lý tài chính

USING WEKA TO DETERMINE CRITERIA IN ASSESSING THE PERFORMANCE OF VIETNAM'S REAL ESTATES ENTERPRISES

Nguyễn Thị Tuyết Nhu, Trần San Đào, Lý Phát Cường, Phan Minh Đạt,
Phan Tấn Anh Phong, Đinh Hùng, Phạm Minh Dũng

ABSTRACT— This research uses WEKA application to select criteria to evaluate the performance of real estate businesses in Vietnam in the period 2015-2019. From WEKA application results, the authors analyze and discuss the reliability of the application results, conclude pros and cons to form lessons for the next WEKA applications.

Nguyễn Thị Tuyết Nhu



Born in 1963
Doctor of Philosophy in
Economics of Viet Nam National
University - Ho Chi Minh City,
2009
English Bachelor degree of Hanoi
University of Foreign Studies
Saigon Languages and Business
College, 1997
nhu.ntt@hufliit.edu.vn

Trần San Đào



Born in 1987
Master of International Business
and Commerce (Finance) – The
University of Sydney, 2012
Bachelor of Economics – Foreign
Trade University – HCMC campus,
2009
dao.ts@hufliit.edu.vn

Lý Phát Cường



Born in 1981
Master of Accounting, Ho Chi
Minh City University of
Technology, 2014
PhD in Accounting, University of
Economics
cuong.lp@hufliit.edu.vn

Phan Minh Đạt



Born in 1987
Master of Professional
Accounting – Central
Queensland University
(Australia), 2012
Bachelor of Enterprise
Finance – University of
Economics in Ho Chi Minh
City, 2009

Phan Anh Tấn Phong



Born in 1974
Master of Sociology,
Department of Social Dynamics,
Graduate School of Humanities,
Kobe University, 2009
Bachelor of Japanese Studies,
Faculty of Oriental Studies, Ho
Chi Minh City University of
Social Sciences, 1998
phongphan@hufliit.edu.vn

Đinh Hùng



Born in 1966
Ph.D Business Administration at
Lac Hong University, 2021
Master Business Administration at
Lincoln University, 2014
Master Computer Science at
University of Science Ho Chi Minh
City, 2003
Bachelor Physics at University of
Science Ho Chi Minh City, 1988
dinhhung621@hufliit.edu.vn

Phạm Minh Dũng



Born in 1978
Master of Computer Science,
University of Science Ho Chi
Minh City, 2012
Ho Chi Minh city University
Foreign Language and
Informatic Technology, 2000
pmdung@hufliit.edu.vn