CLASSIFICATION OF CUSTOMERS BASED ON BEHAVIOR, USING DATA MINING TECHNIQUES

Tran Thi Xuan[1], Nguyen Van Nui[2*]
[1]*TNU - University of Economics and Business Administration*
[2]*TNU - University of Information and Communication Technology*

THÔNG TIN BÀI BÁO

Received: 08/9/2021
Revised: 09/11/2021
Published: 10/11/2021

TỪ KHÓA

Customer classification
Data mining
CMR
Naïve Bayes Classification
Decision Tree
Random Forest

TÓM TẮT

Data mining (DM) is a popular technique, and has been used to extract useful information from existing data, thereby assisting in making decisions that benefit the future. In this paper, the authors focus on the problem of customer classification, thereby helping to find a group of potential customers using Decision Tree J48, Naïve Bayes Classification and Random Forest. The results show that the model based on the Decision Tree gives highest accuracy and feasibility in predicting customer behavior. This result is expected to be an effective suggestion for an approach that can effectively help researchers related to finding a group of potential customers in the banking field.

PHÂN LỚP KHÁCH HÀNG DỰA TRÊN HÀNH VI, SỬ DỤNG KỸ THUẬT KHAI PHÁ DỮ LIỆU

Trần Thị Xuân[1], Nguyễn Văn Núi[2*]
[1]*Trường Đại học Kinh tế và Quản trị kinh doanh – ĐH Thái Nguyên*
[2]*Trường Đại học Công nghệ Thông tin và Truyền thông – ĐH Thái Nguyên*

ARTICLE INFO

Ngày nhận bài: 08/9/2021
Ngày hoàn thiện: 09/11/2021
Ngày đăng: 10/11/2021

KEYWORDS

Phân lớp khách hàng
Khai phá dữ liệu
CRM
Naïve Bayes Classification
Decision Tree
Random Forest

ABSTRACT

Khai phá dữ liệu là một kỹ thuật phổ biến, được sử dụng để trích xuất thông tin hữu ích từ dữ liệu đã có, từ đó hỗ trợ ra các quyết định có lợi cho tương lai. Trong bài báo này, nhóm tác giả tập trung vào vấn đề phân lớp khách hàng, từ đó hỗ trợ tìm ra nhóm khách hàng tiềm năng bằng phương pháp cây quyết định Decision Tree J48, Naïve Bayes Classification và rừng ngẫu nhiên Random Forest. Kết quả cho thấy, mô hình dựa trên thuật toán cây quyết định cho độ chính xác cao nhất, có tính khả thi cao trong việc phân lớp dự đoán hành vi khách hàng. Kết quả này được kỳ vọng sẽ là gợi ý hiệu quả về một hướng tiếp cận cho các nhà phân tích khách hàng trong việc tìm ra nhóm khách hàng tiềm năng thuộc lĩnh vực ngân hàng.

DOI: https://doi.org/10.34238/tnu-jst.4954

* Corresponding author. *Email: nvnui@ictu.edu.vn*

1. Giới thiệu chung

Khai phá dữ liệu là một trong những lĩnh vực nghiên cứu quan trọng và ngày càng phát triển với mục đích trích xuất thông tin từ số lượng lớn các tập dữ liệu tích lũy. Trong thời đại hiện nay, khai phá dữ liệu trở nên phổ biến trong lĩnh vực ngân hàng và là phương pháp phân tích hiệu quả cho phát hiện thông tin hữu ích và chưa biết trong dữ liệu ngân hàng [1]-[3].

Nhận diện khách hàng tiềm năng là công việc đầu tiên trong quá trình quản lý quan hệ khách hàng (Customer Relationship Management - CRM), bao gồm các công việc chính là phân loại và phân tích khách hàng. Khách hàng được chia thành các tập con nhỏ hơn với các thuộc tính giống nhau. Mục tiêu của phân loại khách hàng là nhằm xác định xem ai là người chắc chắn sẽ mua sản phẩm/ dịch vụ. Khai phá dữ liệu (Data mining) được sử dụng phổ biến trong giai đoạn này để hỗ trợ việc nhận diện khách hàng tiềm năng.

Phân loại khách hàng và hệ tư vấn, khuyến nghị khách hàng tín dụng, phát hiện và cảnh báo rủi ro là bước quan trọng trong việc tìm kiếm những khách hàng tiềm năng của ngân hàng. Để thực hiện được việc đó, các nghiên cứu đã thực hiện trên các thuật toán khai phá dữ liệu khác nhau để tìm ra lời giải cho bài toán của mình. Khách hàng được phân loại bằng các thuật toán phân loại trong các kỹ thuật khai phá dữ liệu. Từ đó tìm ra được nhóm khách hàng có cùng sở thích sử dụng các dịch vụ, tiếp sau đó ngân hàng sẽ có chiến lược riêng cho từng nhóm khách hàng như vậy.

Trong những năm gần đây, kỹ thuật khai phá dữ liệu và phân lớp đã được áp dụng thành công trong việc đề xuất mô hình hỗ trợ khác nhau để nâng cao chất lượng dịch vụ [4]-[10].

Nhóm tác giả Sheel Singhal và Dr. G.N. Singh [4] đã đề xuất phương pháp phân lớp bằng việc khai phá luật kết hợp CBA (Classification Based Association rules) để tìm ra các dịch vụ ngân hàng mà khách hàng thường hay sử dụng kèm với một dịch vụ ngân hàng khác. Trong một nghiên cứu khác của Ikizer và cộng sự [5], mạng nơ-ron và các kỹ thuật truyền thống đã phân tích, áp dụng để xây dựng xếp hạng mô hình cho công đoàn vay vốn. Trong nghiên cứu này, Ikizer và cộng sự của mình đã sử dụng mẫu dữ liệu nhất quán bao gồm 18 thuộc tính về ba hiệp hội tín dụng và nghiên cứu của ông đã chứng minh rằng, mạng nơ-ron nhân tạo hữu ích hơn trong dự báo các khoản vay khó đòi, trong khi hồi quy logistic hữu ích trong việc phát hiện các khoản nợ xấu và tốt với tỉ lệ dự đoán chính xác 77%.

Do vai trò rất quan trọng trong việc phân lớp nhận diện khách hàng tiềm năng, số lượng nghiên cứu để tìm hiểu sâu rộng về vấn đề này đã tăng nhanh trong những năm qua. Gần đây, có một vài mô hình phân lớp được nghiên cứu, đề xuất để hỗ trợ các nhà nghiên cứu trong việc phân lớp, dự đoán khách hàng tiềm năng [4]-[10]. Tuy nhiên, ở thời điểm hiện tại, vẫn còn thiếu các mô hình tính toán phù hợp và công cụ dự đoán với độ chính xác cao có thể hỗ trợ hiệu quả cho việc phân loại nhận diện khách hàng, đặc biệt là nhận diện nhóm khách hàng tiềm năng thuộc lĩnh vực ngân hàng. Bên cạnh đó, do sự tiến bộ của khoa học kỹ thuật và ảnh hưởng của cách mạng công nghiệp 4.0, dữ liệu khách hàng đã kiểm chứng thực nghiệm đang ngày càng được bổ sung nhiều hơn. Chính vì vậy, việc thiếu hụt mô hình phân lớp phân loại khách hàng là một vấn đề cấp thiết cần được quan tâm giải quyết.

Tiếp tục phát triển các ý tưởng nghiên cứu trước đây, trong bài viết này nhóm tác giả tập trung vào vấn đề phân lớp khách hàng hỗ trợ tìm ra nhóm khách hàng tiềm năng bằng phương pháp cây quyết định J48, Naive Bayes và rừng ngẫu nhiên.

2. Xây dựng, huấn luyện mô hình

2.1. Thu thập, tiền xử lý dữ liệu

Trong nghiên cứu này, bộ dữ liệu đã kiểm chứng thực nghiệm từ nghiên cứu của nhóm tác giả Moro và cộng sự [1], [2] được lựa chọn sử dụng để xây dựng và huấn luyện mô hình. Bộ dữ liệu sử dụng cho nghiên cứu này được thu thập từ kho dữ liệu học máy UCI [11], bao gồm thông tin

của 45211 khách hàng (từ tháng 5 năm 2008 đến tháng 11 năm 2010) với 17 thuộc tính được thể hiện chi tiết ở Bảng 1.

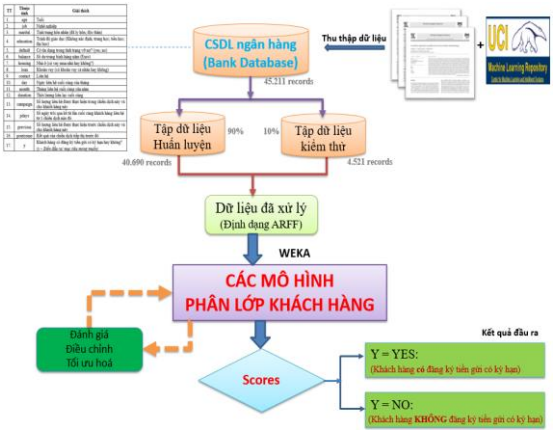
Bảng 1. *Thông tin bộ dữ liệu khách hàng sử dụng trong nghiên cứu này*

TT	Thuộc tính	Giải thích
1	age	Tuổi
2	job	Nghề nghiệp
3	marital	Tình trạng hôn nhân (đã ly hôn; độc thân)
4	education	Trình độ giáo dục (Không xác định; trung học; tiểu học; đại học)
5	default	Có tín dụng trong tình trạng vỡ nợ? (yes; no)
6	balance	Số dư trung bình hàng năm (Euro)
7	housing	Nhà ở (có vay mua nhà hay không?)
8	loan	Khoản vay (có khoản vay cá nhân hay không)
9	contact	Liên hệ
10	day	Ngày liên hệ cuối cùng của tháng
11	month	Tháng liên hệ cuối cùng của năm
12	duration	Thời lượng liên lạc cuối cùng
13	campaign	Số lượng liên hệ được thực hiện trong chiến dịch này và cho khách hàng này
14	pdays	Số ngày trôi qua kể từ lần cuối cùng khách hàng liên hệ từ 1 chiến dịch nào đó
15	previous	Số lượng liên hệ được thực hiện trước chiến dịch này và cho khách hàng này
16	poutcome	Kết quả của chiến dịch tiếp thị trước đó
17	**y**	Khách hàng có đăng ký tiền gửi có kỳ hạn hay không? *(y – Biến đầu ra/ mục tiêu mong muốn)* *y ="yes":* Khách hàng **có** mở tài khoản tiết kiệm có kỳ hạn *y ="no":* Khách hàng **không** mở tài khoản tiết kiệm có kỳ hạn

Để xây dựng dữ liệu huấn luyện (training data) và dữ liệu kiểm thử (testing data), trong nghiên cứu này, chúng tôi tiến hành lấy ngẫu nhiên 10% khách hàng từ tổng số 45.211 khách hàng đã thu được trước đó làm dữ liệu kiểm thử. Phần còn lại gồm 90% khách hàng sẽ được sử dụng để xây dựng dữ liệu huấn luyện.

2.2. Xây dựng và huấn luyện mô hình

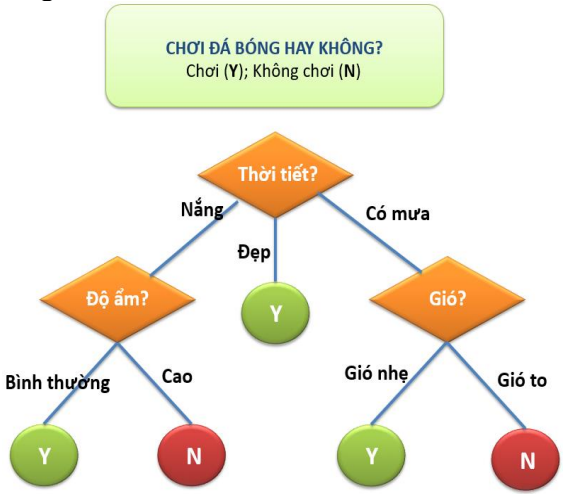
Trong bài báo này, mô hình phân lớp khách hàng được xây dựng và huấn luyện dựa trên hành vi của khách hàng trong lĩnh vực ngân hàng, sử dụng một số kỹ thuật khai phá dữ liệu gồm Naïve Bayes, Decision Tree và Random Forest [3]. Mô hình tổng thể phân lớp khách hàng đề xuất trong bài báo này được thể hiện chi tiết ở Hình 1.



Hình 1. *Sơ đồ hệ thống phân lớp khách hàng*

Decision Tree (Cây quyết định) là một mô hình học máy thuộc nhóm thuật toán học có giám sát (supervised learning). Nó là một phương pháp học máy mạnh và phổ biến đã được biết đến và áp dụng thành công cho bài toán khai phá dữ liệu và phân lớp. Cây quyết định chính là cây mà mỗi nút biểu diễn một đặc trưng, mỗi nhánh (branch) biểu diễn một quy luật (rule), mỗi nút lá biểu diễn một kết quả (giá trị cụ thể hoặc một nhánh tiếp tục). Cây quyết định có thể được dùng cho bài toán phân lớp dữ liệu bằng cách xuất phát từ gốc của cây và di chuyển theo các nhánh cho đến khi gặp nút lá.

Một ví dụ về cây quyết định được mô tả nguyên tắc (luật) để quyết định **CHƠI (Y)** hay **KHÔNG CHƠI (N)** đá bóng của các bạn nam sinh viên được minh họa như ở Hình 2.



Hình 2. *Cây quyết định về việc Choi (Y) hay Không chơi (N) đá bóng của các bạn nam sinh viên*

Dựa theo mô hình cây quyết định ở Hình 2, ta có thể thấy được quy tắc để biết các bạn nam sinh viên quyết định có đi chơi đá bóng hay không (dựa trên các thông tin liên quan đến thời tiết, độ ẩm, gió) sẽ như sau:

**** Chơi đá bóng (Y) nếu thoả mãn 1 trong các điều kiện sau:***

(1) Thời tiết đẹp
(2) Trời nắng, độ ẩm bình thường
(3) Trời có mưa, gió nhẹ

**** Không chơi đá bóng (N) nếu:***

(1) Trời nắng, độ ẩm cao
(2) Trời mưa, gió to

Naïve Bayes Classification (NBC) là một thuật toán dựa trên định lý Bayes về lý thuyết xác suất để đưa ra các phán đoán cũng như phân loại dữ liệu dựa trên các dữ liệu được quan sát và thống kê. NBC là một trong những thuật toán được ứng dụng rất nhiều trong các lĩnh vực Machine learning dùng để đưa các dự đoán chính xác nhất dựa trên một tập dữ liệu đã được thu thập, vì nó khá dễ hiểu và độ chính xác cao. Nó thuộc vào nhóm Supervised Machine Learning Algorithms (thuật toán học có hướng dẫn), tức là máy học từ các ví dụ từ các mẫu dữ liệu đã có.

Công thức của định luật Bayes được phát biểu như sau:

$$P(A|B) = \frac{P(B|A)X\ P(A)}{P(B)}$$

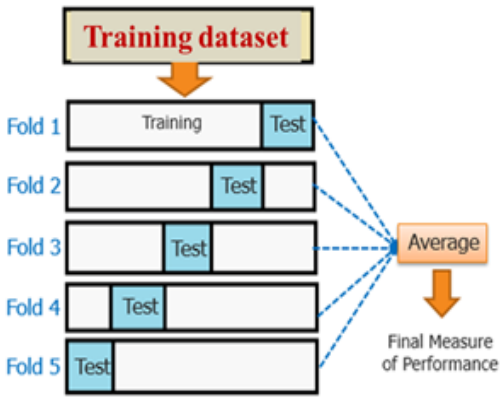
Trong đó:

- P(A|B) là xác suất xảy ra của một sự kiện ngẫu nhiên A khi biết sự kiện liên quan B đã xảy ra.
- P(B|A) là xác suất xảy ra B khi biết A xảy ra.
- P(A) là xác suất xảy ra của riêng A mà không quan tâm đến B.
- P(B) là xác suất xảy ra của riêng B mà không quan tâm đến A.

Random Forest (RF) là thuật toán học có giám sát (supervised learning). RF có thể được sử dụng cho cả phân lớp và hồi quy. RF cũng là thuật toán linh hoạt và dễ sử dụng nhất. Một khu rừng bao gồm cây cối. Người ta nói rằng càng có nhiều cây thì rừng càng mạnh. Random forests tạo ra cây quyết định trên các mẫu dữ liệu được chọn ngẫu nhiên, được dự đoán từ mỗi cây và chọn giải pháp tốt nhất bằng cách bỏ phiếu.

Với bài toán phân lớp: cho một tập dữ liệu huấn luyện $D = \{(d_i)\}_{i=1}^{N} = \{(x_i, y_i)\}_{i=1}^{N}$ với x_i là vector M chiều, $y_i \in Y$, trong đó: Y gọi là lớp, giả sử có C nhãn lớp $Y \in \{1, 2, \ldots, C\}(C \geq 2)$. Ý tưởng chính của mô hình Random forest là lựa chọn ngẫu nhiên 2 lần (ngẫu nhiên mẫu và ngẫu nhiên thuộc tính) trong suốt quá trình xây dựng cây.

Để đánh giá hiệu năng của mô hình, 2 phương pháp phổ biến được sử dụng đó là: đánh giá chéo 5-mặt (5-fold cross-validation) và kiểm thử độc lập (Independent testing) sử dụng bộ dữ liệu độc lập (independent testing dataset với bộ dữ liệu huấn luyện (training dataset). Với phương pháp đánh giá chéo 5 mặt (Như hiển thị ở Hình 3, tập dữ liệu huấn luyện sẽ được chia ngẫu nhiên thành 5 tập con bằng nhau, lần lượt mỗi tập con sẽ được dùng cho vai trò kiểm thử trong khi 4 tập còn lại được dùng làm dữ liệu huấn luyện.



Hình 3. *Mô hình kiểm tra đánh giá chéo 5-mặt*

Các đại lượng thông dụng được sử dụng để đo lường và đánh giá hiệu năng của mô hình bao gồm: Accuray (độ chính xác), MCC (hệ số tương quan Matthews và Error Rate [6]-[11].

$$ACC = \frac{TP+TN}{P+N}; \qquad Error\ Rate = \frac{FP+FN}{P+N}$$

$$MCC = \frac{(TP \times TN) - (FN \times FP)}{\sqrt{(TP + FN) \times (TN + FP)(TP + FP)(TN + FN)}}$$

Trong đó:

P: Số bản ghi Positive trong tập dữ liệu

N: Số bản ghi Negative trong tập dữ liệu

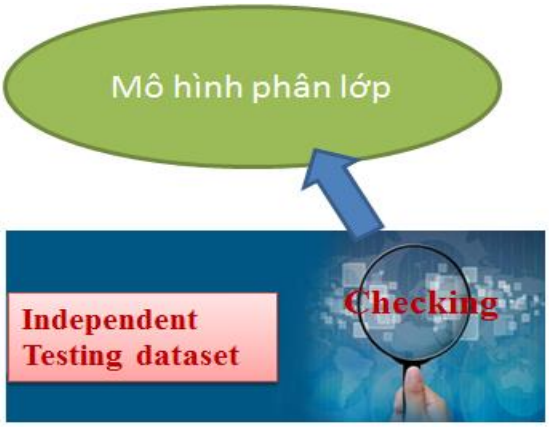
TP: Số bản ghi Positive (y = ”yes”) được dự đoán là Positive

TN: Số bản ghi Negative (y = ”no”) được dự đoán là Negative.

FP: Số bản ghi Negative (y = ”no”) được dự đoán là Positive

FN: Số bản ghi Positive (y = ”yes”) được dự đoán là Negative.

Ngoài ra, phương pháp kiểm thử, đánh giá độc lập cũng được sử dụng để đánh giá hiệu năng của mô hình phân lớp, dự đoán. Như hiển thị ở Hình 4, theo phương pháp đánh giá kiểm thử độc lập, hiệu năng của mô hình sẽ được xác định bằng việc sử dụng một bộ dữ liệu kiểm thử hoàn toàn khác biệt và không trùng lặp với bộ dữ liệu huấn luyện đã dùng cho việc huấn luyện mô hình (Independent testing dataset). Việc sử dụng bộ dữ liệu kiểm thử độc lập này sẽ giúp ta kiểm tra, đánh giá một cách khách quan nhất hiệu năng phân lớp của mô hình.



Hình 4. *Mô hình kiểm thử độc lập*

3. Kết quả và một số thảo luận

3.1. Kết quả huấn luyện và đánh giá mô hình phân lớp theo phương pháp đánh giá chéo 5-mặt

Như đã trình bày trước đó, trong nghiên cứu này, chúng tôi tiến hành sử dụng các thuật toán khai phá dữ liệu như NBC, RF, J48 để xây dựng và huấn luyện mô hình phân lớp dự đoán khách hàng có mở tài khoản tiết kiệm có kì hạn hay không. Theo thông tin tổng hợp ở Bảng 2, mô hình đạt hiệu năng phân lớp với độ chính xác của thuật toán Decision Tree J48 là 90,46%, giá trị MCC = 0,497.

Bảng 2. *Kết quả đánh giá mô hình bằng phương pháp đánh giá chéo 5-mặt*

Model	ACC	SEN	SPE	MCC
NBC	87,98%	52,6%	92,6%	0,437
RF	90,39%	63,3%	96,7%	0,469
J48	90,46%	49,00%	96,00%	0,497

3.2. Kết quả đánh giá mô hình sử dụng phương pháp kiểm thử độc lập

Như đã đề cập trước đó, phương pháp đánh giá độc lập giúp kiểm chứng khả năng thực nghiệm của mô hình trong trường hợp thực tế, khách quan nhất. Để thực hiện được việc này, một bộ dữ liệu kiểm thử độc lập đã được xây dựng bao gồm 521 dữ liệu positive và 4000 dữ liệu negative.

Kết quả kiểm tra đánh giá hiệu năng của mô hình khi tiến hành bởi phương pháp kiểm thử độc lập được thể hiện chi tiết ở Bảng 3. Qua các con số thể hiện ở Bảng 3, ta thấy rằng mô hình đạt độ chính xác tương đối cao và có tính khả thi tốt trong việc dự đoán quyết định mở tài khoản tiết kiệm có kì hạn của khách hàng. Ở phương pháp này, mô hình dự đoán độ chính xác cao nhất sử dụng thuật toán rừng ngẫu nhiên RF cho kết quả cao nhất với độ chính xác là 90,44% với MCC = 0,501.

Bảng 3. *Kết quả đánh giá mô hình bằng phương pháp kiểm thử độc lập*

Model	ACC	SEN	SPE	MCC
NBC	88,05%	51,5%	93,2%	0,447
RF	90,44%	52,5%	96,8%	0,501
J48	90,29%	49,3%	95,9%	0,468

Để minh hoạ thêm cho hiệu quả của mô hình đề xuất trong việc dự đoán hành vi khách hàng, từ đó tìm kiếm khách hàng tiềm năng cho lĩnh vực ngân hàng; chúng thôi xin đưa ra một số kết quả thu được từ thuật toán NBC như thể hiện ở Bảng 4. Theo thông tin từ Bảng 4**,** liên quan đến nghề nghiệp của khách hàng thì nhóm doanh nhân (Entrepreneur) là nhóm khách hàng tiềm năng nhất cho quyết định mở tài khoản tiết kiệm có kỳ hạn. Tương tự, nhóm khách hàng chưa có nhà ở, nhóm khách hàng chưa có gia đình (hoặc đã ly hôn) cũng sẽ là nhóm khách hàng tiềm năng nhất cho quyết định mở tài khoản tiết kiệm có kỳ hạn.

Bảng 4. *Kết quả thu được từ thuật toán NBC*

Thuộc tính	Class Y = NO	Class Y = Yes	Tỉ lệ có quyết định Y = Yes
Job			
Management	8158	1302	13,76%
Technician	6758	841	11,07%
Entrepreneur	**1365**	**709**	**34,19%**
Blue – Collar	9025	709	7,28%
Unknown	255	35	12,07%
Retired	1749	517	22,82%
Admin	4541	632	12,22%
Services	3786	370	8,90%
Self-employed	1102	203	15,56%
Housemaid	1132	110	8,86%
Student	670	270	28,72%
Housing			
Yes	23196	1936	7,70%
No	**16728**	**3355**	**16,71%**
Marital			
Married	24460	2756	10,13%
Singer	**10879**	**1913**	**14,95%**
Divorced	4586	623	11,96%

4. Kết luận

Qua kết quả phân lớp trên, ta thấy rằng cả 3 mô hình phân lớp khách hàng đều đạt độ chính xác đến 90%, trong đó mô hình phân lớp dựa trên thuật toán cây quyết định cho kết quả cao nhất. Điều này cho thấy các mô hình phân lớp ở trên, đặc biệt là thuật toán dựa trên cây quyết định rất phù hợp với bài toán phân lớp dự đoán khách hàng thuộc lĩnh vực ngân hàng.

Ngoài ra, thông qua các kết quả nhận được từ một số mô hình phân lớp ở trên, đặc biệt là mô hình phân lớp dựa vào thuật toán Naïve Bayes NBC ở Bảng 4 ta có thể biết được một khách hàng có đặc điểm gì thì sẽ là khách hàng tiềm năng.

Theo tiêu chí nghề nghiệp khách hàng thuộc nhóm doanh nhân (Entrepreneur) mở tài khoản tiết kiệm có kì hạn nhiều nhất. Dựa theo tiêu chí Housing, khách chưa có nhà có xu hướng mở tài khoản tiết kiệm có kì hạn nhiều hơn nhóm khách hàng đã sở hữu nhà ở. Dựa theo tiêu chí kết hôn, tỉ lệ khách hàng chưa kết hôn và đã ly hôn mở tài khoản tiết kiệm nhiều hơn nhóm người đã kết hôn.

Từ phân tích trên ta thấy, khách hàng doanh nhân, khách hàng chưa có nhà, khách hàng độc thân và đã ly hôn là những khách hàng tiềm năng, cần khai thác thêm những khách hàng có đặc điểm như trên để tư vấn, thuyết phục hay có những chính sách để khách hàng trở thành khách hàng tiềm năng.

TÀI LIỆU THAM KHẢO/ REFERENCES

[1] S. Moro, R. Laureano, and P. Cortez, "Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology," In P. Novais et al. (Eds.), *Proceedings of the European Simulation and Modelling Conference - ESM'2011*, Guimaraes, Portugal, October, 2011, pp. 117-121.

[2] S. Moro, P. Cortez, and P. Rita, "A Data-Driven Approach to Predict the Success of Bank Telemarketing," *Decision Support Systems, Elsevier*, vol. 62, pp. 22-31, June 2014.

[3] V. L. M. E. Oliveira, "Analytical Customer Relationship Management in Retailing Supported by Data Mining Techniques," PhD, Industrial Engineering and Management, Universidade do Porto, 1, 2019.

[4] S. Singhal and G. N. Singh, "Classification using Association Rule Mining," *International Journal of Computer Sicence & Communication*, vol. 3, no. 2, pp. 256-258, 2012.

[5] İ. Nazlı and H. A. Guvenir. "Mining interesting rules in bank loans data," *Proceedings of the Tenth Turkish Symposium on Artificial Intelligence and Neural Networks*, 2001.

[6] F. Akhyani and A. Komeili, *New approach based on proximity/remoteness measurement for customer classification*, Electronic Comerce Research Springer, 2020.
[7] A. Suyanto, "Developing an LSTM-based Classification Model of IndiHome Customer Feedbacks," International Conference on Data Science and Its Applications (ICoDSA), Indonesia, 2020.
[8] H. Y. Lam and Y. P. Tsang, *Data analytics and the P2P cloud: an integrated model for strategy formulation based on custommer behaviour*, Springer, 2020.
[9] A. J. Hamid and T. M. Ahmed, "Developing Prediction Model of Loan Risk in Banks Using Data Mining," *Machine Learning and Applications*, vol. 3, p. 9, 2016.
[10] D. Tomar and S. Agarwal, "A survey on Data Mining approaches for Healthcare," *International Journal of Bio-Science and Bio-Technology*, vol. 5, pp. 241-266, 2013.
[11] D. Dua and C. Graff, "UCI Machine Learning Repository," Irvine, CA: University of California, School of Information and Computer Science, 2019. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/bank+marketing. [Accessed June 20, 2021].