

## TỔNG QUAN

**Tín dụng xoay vòng** cho phép người vay rút tiền nhiều lần trong hạn mức tín dụng, trả lãi trên số tiền đã rút, thay vì nhận một khoản tiền lớn một lần như vay theo hình thức truyền thống. Điều này giúp doanh nghiệp và cá nhân quản lý dòng tiền **linh hoạt** hơn, đáp ứng nhu cầu tài chính ngắn hạn không thể dự đoán trước. Để đưa ra hợp đồng tín dụng xoay vòng hợp lý, ngân hàng cần quản lý **rủi ro** từ lạm phát, lãi suất và hành vi khách hàng, đồng thời đảm bảo lợi nhuận đủ để hoạt động.

Để có cái nhìn đơn giản và tổng quan, trước tiên bài báo cáo trả lời các câu hỏi đề bài đã đặt ra trong một vài trường hợp về giá trị tiền theo thời gian, giá trị của các lần rút với mức lãi suất cụ thể. Sau đó, bài báo cáo sẽ xây dựng các mô hình để mô phỏng và dự đoán hành vi của người vay bằng hàm tính toán **đệ quy**. Từ đó, dựa vào hành vi của khách hàng mà chúng tôi xây dựng được mô hình dự đoán quy luật rút tiền và thời gian đáo hạn của người vay bằng thuật toán được xây dựng trên **cây quyết định hồi quy (Decision Tree Regressor)** và xây dựng một thuật toán cải tiến từ thuật toán **Random Forest Regressor** với độ chính xác cao hơn bằng ý tưởng chia để trị.

Kết quả chúng tôi tạo ra một mô hình toán có thể định giá hợp đồng tín dụng xoay vòng, giúp đánh giá lợi ích mà công ty B có thể thu được dựa trên mức lãi suất hợp đồng. Điều này giúp ngân hàng thiết lập mức lãi suất hợp lý nhất và áp dụng vào thực tế. Mô hình này được đánh giá là tối ưu về mặt toán học và mang lại lợi ích đáng kể cho cả hai bên. Hơn nữa, tính **linh hoạt** và không bị ràng buộc bởi dữ liệu cụ thể giúp mô hình của chúng tôi có thể được áp dụng trong nhiều ngữ cảnh khác nhau.

**Từ khóa:** Tín dụng xoay vòng, linh hoạt, rủi ro, đệ quy, Decision Tree Regressor, cây quyết định (Decision Tree), Random Forest Regressor.

MỤC LỤC

TỔNG QUAN.....	1
MỤC LỤC.....	2
I. ĐẶT VẤN ĐỀ.....	3
II. MÔ HÌNH HÓA BÀI TOÁN VÀ CÁC GIẢ SỬ.....	3
III. HƯỚNG GIẢI QUYẾT.....	3
1. Giá trị thời gian của tiền.....	3
1.1. Vấn đề 1.1.....	3
1.2. Vấn đề 1.2.....	4
2. Rủi ro lãi suất.....	4
2.1. Vấn đề 2.1.....	4
2.2. Vấn đề 2.2.....	4
2.3. Vấn đề 2.3.....	5
3. Mô hình hoá quy luật rút tiền.....	6
4. Mô hình mô phỏng sự biến động của lãi suất và hành vi khách hàng.....	8
IV. ĐÁNH GIÁ MÔ HÌNH.....	13
1. Nhận xét chung về bài toán.....	13
2. Mô hình thứ nhất.....	14
3. Mô hình thứ hai.....	14
4. So sánh 2 mô hình.....	14
V. TÍNH ỨNG DỤNG CỦA MÔ HÌNH BÀI TOÁN.....	14
VI. TRÍCH DẪN - TÀI LIỆU THAM KHẢO.....	15
VII. PHỤ LỤC.....	15
VIII. LỜI CẢM ƠN.....	21

## I. ĐẶT VẤN ĐỀ

Hình thức tín dụng này đặc biệt phổ biến trong lĩnh vực kinh doanh và cá nhân vì tính linh hoạt và tiện lợi của nó, giúp công ty và cá nhân quản lý dòng tiền hiệu quả hơn, rất hữu ích cho việc đáp ứng các nhu cầu tài chính ngắn hạn và không dự đoán trước được. Qua đó, công ty có thể tập trung vào kinh doanh và đóng góp một phần không nhỏ cho GDP và sự phát triển kinh tế của đất nước/ thế giới. Tín dụng xoay vòng giúp công ty đi vay giảm thiểu rủi ro trong việc quản lý nguồn vốn và giao trách nhiệm đó cho phía ngân hàng, những người có chuyên môn cao hơn. Vấn đề đặt ra là ngân hàng phải đưa ra một lãi suất hợp đồng phù hợp để thu được lợi nhuận đủ để đáp ứng nhu cầu hoạt động trước những rủi ro và biến động khó lường của thị trường như giá trị thời gian của tiền, lãi suất thị trường và hành vi khách hàng, cũng như thích hợp để công ty vay sẵn sàng ký vào hợp đồng. Để có thể đưa ra một hợp đồng tín dụng xoay vòng hợp lý, đáp ứng nhu cầu của khách hàng cũng như ngân hàng, ngân hàng cần giải quyết tất cả các rủi ro và biến động, và điều này đã trở thành một vấn đề nan giải của cơ quan chính phủ trong việc quản lý các ngân hàng.

Nhận thấy yêu cầu cấp thiết đó, nhóm chúng tôi đã suy nghĩ về việc xây dựng một mô hình dùng để quy luật hóa các thời điểm rút tiền, đáo hạn và số tiền mỗi lần rút của mỗi “bullet” để định giá hợp đồng. Ý tưởng đó đã dẫn đến bài toán và mô hình đã nêu. Tuy nhiên, do số liệu và thời gian có hạn, ở đây chúng tôi chỉ xét một khoảng thời gian đủ lâu và mang tính tổng quan nhất. Nếu mở rộng theo cách này trên một quy mô lớn hơn, chúng ta có thể giải quyết cho nhu cầu lớn hơn nhiều so với mô hình.

## II. MÔ HÌNH HÓA BÀI TOÁN VÀ CÁC GIẢ SỬ

Do có một vài điều kiện phức tạp ảnh hưởng đến bài toán và để xây dựng mô hình được thuận tiện hơn, báo cáo này đã đơn giản hóa bằng một số giả sử. Tuy nhiên, do các vấn đề của bài toán tổng được chia thành các vấn đề nhỏ hơn, nên chúng tôi sẽ trình bày chi tiết các giả sử trong từng vấn đề.

## III. HƯỚNG GIẢI QUYẾT

### 1. Giá trị thời gian của tiền

Bài toán: Giả sử công ty A và ngân hàng B ký một hợp đồng tín dụng xoay vòng ở thời điểm hiện tại, tháng 5 năm 2024 với số tiền là 1000 tỷ VNĐ, lãi suất của hợp đồng là 3.5% với kỳ hạn 10 năm. Lãi suất thị trường hiện tại là 2.9%. Một năm sau, tháng 5 năm 2025, công ty A rút 100 tỷ VND với thời hạn 1 năm. Tuy nhiên lãi suất của thị trường ở thời điểm đó đã là 3.75%, như vậy ngân hàng B phải huy động vốn trên thị trường với lãi suất 3.75% để cho công ty A vay lại với lãi suất 3.5% như cam kết trong hợp đồng.

1.1. Hãy tính giá trị của lần rút tiền này đối với ngân hàng B (ở thời điểm hiện tại)?

1.2. Trong câu hỏi trên, lãi suất của hợp đồng phải là bao nhiêu để cả bên A và bên B đều sẵn sàng ký vào hợp đồng?

#### 1.1. Vấn đề 1.1

Gọi:

- $r$ : lãi suất thị trường hiện tại (2.9%)
- $T$ : số năm (năm)
- $e^{-rT}$ : giá trị 1 đồng ở thời điểm  $T$
- Theo giả thiết: 1 đồng ở thời điểm  $T$  năm sẽ có giá trị là  $e^{-rT}$  ở thời điểm hiện tại (giá trị của tiền theo thời gian).

Thay  $T = 1$  và  $r = 0.029$  ( $= 2.9\%$ ), chúng ta có 1 đồng ở thời điểm 05/2025 tương ứng với số tiền ở thời điểm 05/2024 là:

$$e^{-0.029 \times 1} \approx 0.9714 \text{ (đồng)}$$

Vậy giá trị của lần rút tiền này đối với ngân hàng B (ở thời điểm hiện tại) là:

$$e^{-0.029 \times 1} \times 1 \approx 0.9714 \text{ (trăm tỷ đồng)}$$

## 1.2. Vấn đề 1.2

Giả sử:

- Công ty A khi trả tiền cho ngân hàng B thì cần trả hết 1 lần số tiền đã vay, đảm bảo sự đồng nhất về giá trị của tiền qua các thời điểm.
- Công ty A sẽ vay tiền tại những thời điểm mà lãi suất thị trường cao nhất. Nghĩa là ta chỉ chọn những phân tích những tình huống bất lợi nhất cho ngân hàng B, những trường hợp khác đơn giản hơn sẽ xử lý tương tự.

Với những điều giả sử ở trên, ta nhận thấy rằng ngân hàng B cần đưa ra lãi suất ít nhất là  $a$  - lãi suất cao nhất trong 10 năm. Vì như vậy, chúng ta sẽ đảm bảo ngân hàng B không lỗ, tuy nhiên điều này sẽ hạn chế quyền lợi của công ty A. Để đảm bảo quyền lợi của hai một cách chính xác, ta cần xét yếu tố về xác suất định tính lãi suất thị trường trong vòng 10 năm sẽ được trình bày rõ hơn ở vấn đề 2.2.

## 2. Rủi ro lãi suất

Bài toán: Trong phần trên, chúng ta giả sử lãi suất thị trường trong một năm tới là 3.75%. Tuy nhiên lãi suất của thị trường ở thời điểm tháng 5 năm 2025 là một đại lượng ngẫu nhiên trong tương lai mà ngân hàng B không biết trước được. Thông qua những nghiên cứu về thị trường, họ chỉ có thể biết là, trong năm tiếp theo:

- $p_{\text{up}} = 25\%$ : Lãi suất của thị trường sẽ tăng lên 1%
- $p_{\text{down}} = 25\%$ : Lãi suất của thị trường sẽ giảm xuống 0.5%
- $p = 1 - p_{\text{up}} - p_{\text{down}} = 50\%$  thì lãi suất của thị trường sẽ vẫn là 2.9%.

2.1. Hãy tính giá trị của lần rút tiền (ở thời điểm hiện tại) với những nhận định về thị trường của ngân hàng B như trên?

2.2. Trong trường hợp này, lãi suất của hợp đồng phải là bao nhiêu để cả bên A và bên B đều sẵn sàng ký vào hợp đồng?

2.3. Hãy sử dụng dữ liệu quá khứ về lãi suất của thị trường để ước lượng  $p_{\text{up}}$  và  $p_{\text{down}}$ . Sau đó định giá bullet theo tham số ước lượng được

### 2.1. Vấn đề 2.1

Tương tự với bài toán 1.1, đầu tiên, chúng tôi sẽ tính giá trị kỳ vọng. Đáp số cần tìm chính là tích của giá trị kỳ vọng với  $e^{-0.029 \times 1}$ . Theo đề bài,  $p_{\text{up}} = 25\%$ : lãi suất của thị trường sẽ tăng lên 1% nên số tiền tính theo lãi suất này là:

$$25\% \times 100 \times (1 + 3.9\%) = 25.975 \text{ (tỷ đồng)}.$$

Ta làm tương tự rồi cộng lại sẽ tính được giá trị kỳ vọng là:

$$25.975 + 50\% \times 100 \times (1 + 2.9\%) + 25\% \times 100 \times (1 + 2.4\%) = 103.025 \text{ (tỷ đồng)}.$$

Giá trị của lần rút tiền (ở thời điểm hiện tại) với những nhận định về thị trường của ngân hàng B là:

$$e^{-0.029 \times 1} \times 103.025 \approx 100.08 \text{ (tỷ đồng)}$$

### 2.2. Vấn đề 2.2

Để có thể giải quyết được vấn đề lãi suất hợp đồng bao nhiêu là phù hợp cho công ty A và công ty B, thỏa mãn 2 điều kiện. Trong đó, điều kiện thứ nhất cần được ưu tiên hơn.:

- Ngân hàng B đảm bảo không bị lỗ.
- Công ty A nhận được lãi suất thấp nhất.

Trước tiên, ta sẽ định nghĩa giá trị kỳ vọng:

- Giá trị kỳ vọng của một biến ngẫu nhiên là trung bình có trọng số là xác suất của tất cả các giá trị cụ thể của biến đó.

- Giá trị kỳ vọng có thể được tính bằng công thức:  $EV = \sum_{i=1}^n x_i \times p_i$  ( $i = \overline{1, n}$ ), trong đó
  - +  $x_i$  là các giá trị của biến.
  - +  $p_i$  là khả năng xảy ra của các biến  $x_i$  tương ứng (%).

Ta áp dụng công thức trên để tính được giá trị kỳ vọng dựa trên giá trị lãi suất:

$$EV = (0.25 \times \frac{100(x - 3.9)}{1.029}) + (0.5 \times \frac{100(x - 2.9)}{1.029}) + (0.25 \times \frac{100(x - 2.4)}{1.029})$$

Ở bài toán trên, để đảm bảo 2 điều kiện đã nêu, thì giá trị kỳ vọng EV phải bằng không:

$$\Leftrightarrow (0.25 \times \frac{100(x - 3.9)}{1.029}) + (0.5 \times \frac{100(x - 2.9)}{1.029}) + (0.25 \times \frac{100(x - 2.4)}{1.029}) = 0$$

$$\Leftrightarrow x = 3.025\%$$

Nhưng để kết luận về giá trị lãi suất hợp đồng lý tưởng, ta cần cộng thêm một lượng chênh lệch lãi suất, để đảm bảo mức độ rủi ro của công ty A mà ngân hàng B cần tính:

$$r_{RCF} = r_m + s$$

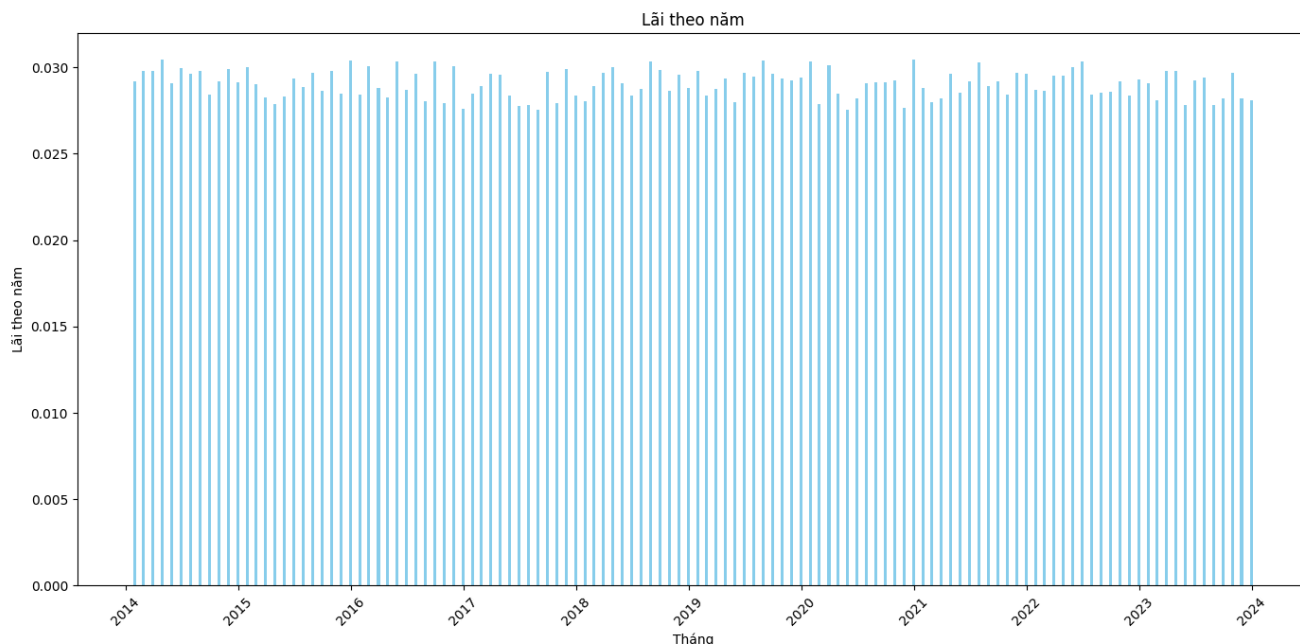
Thay  $r_m = 3.025\%$  và  $s = (3.5\% - 2.9\%) = 0.6\%$ , ta được

$$y = 3.025\% + 0.6\% = 3.625\%$$

Đây chính là lãi suất hợp đồng lý tưởng thỏa mãn điều kiện của bài toán.

### 2.3. Vấn đề 2.3

Trước tiên, để giải quyết vấn đề, chúng tôi đã thu thập được các đầy đủ các số liệu về lãi suất thị trường theo trích dẫn appendix, nhưng ở đây chúng tôi trình bày các số liệu dưới dạng biểu đồ thanh:



Hình 1. Biểu đồ lãi theo năm từ năm 2014 đến 2024

Chúng tôi nhận thấy rằng, các số liệu về lãi suất thị trường qua các năm từ năm 2016 đến năm 2023 dao động xung quanh mức 2.9%. Cụ thể hơn:

- Các số liệu nếu thấp hơn 2.9%, hiệu giữa 2.9% và các số liệu đó sẽ khoảng 0.5%
- Các số liệu nếu cao hơn 2.9%, hiệu giữa các số liệu đó và 2.9% sẽ khoảng 1%

Chính vì vậy, để đơn giản hóa cách tính, chúng tôi sẽ làm tròn các số liệu như sau:

- Các số liệu thấp hơn 2.9% bằng 2.4%.
- Các số liệu cao hơn 2.9% bằng 3.9%

Tiếp theo, chúng ta sẽ thống kê số lượng số lớn hơn 3.9% và số lượng số nhỏ hơn 2.4% để định lượng  $p_{\text{up}}, p_{\text{down}}$ . Từ đó, ta dễ dàng đếm được  $p_{\text{up}} = 25\%$ ,  $p_{\text{down}} = 25\%$  và  $p_{\text{giữ nguyên}} = 50\%$

Từ đó chúng ta cần tính giá trị hiện tại của khoản tiền 100 tỷ mà công ty A sẽ rút vào tháng 5 năm 2025 dựa trên xác suất lãi suất trong tương lai. Theo công thức đã thiết lập cho giá trị kỳ vọng:

$$EV = 0.25 \times \frac{100}{1+0.039} + 0.25 \times \frac{100}{1+0.024} + 0.50 \times \frac{100}{1+0.029} \approx 97.05 \text{ (tỷ VNĐ)}$$

Giả sử lãi suất hợp đồng là  $r$ , giá trị kỳ vọng của khoản tiền 100 tỷ là:

$$\begin{aligned} EV &= \frac{100}{1+r} \\ \Leftrightarrow 1+r &= \frac{100}{97.05} \\ \Leftrightarrow r &= \frac{100}{97.05} - 1 \approx 0.0304 \end{aligned}$$

Vậy lãi suất hợp đồng lý tưởng là 3.04% thì cả 2 bên sẽ đồng ý ký vào hợp đồng. Từ lãi suất hợp đồng lý tưởng ta dễ dàng có được giá trị của các bullet trong hợp đồng vay tín dụng xoay vòng.

### 3. Mô hình hoá quy luật rút tiền

Bài toán: Giả sử công ty A chỉ rút tiền và trả tiền vào cuối tháng. Ta giả sử thêm là lãi suất thị trường sẽ không biến động theo thời gian và mãi được duy trì ở 2.9%. Giá trị của hợp đồng tín dụng xoay vòng bây giờ chỉ phụ thuộc vào quy luật rút tiền của công ty A (hay hành vi của công ty A).

Hãy mô hình hoá quy luật rút tiền của công ty A, bao gồm:

- Thời điểm rút tiền.
- Hoặc thời gian đáo hạn mỗi lần rút.
- Hoặc số tiền rút mỗi lần.
- Hoặc tất cả các biến số trên của mỗi bullet để định giá hợp đồng tín dụng xoay vòng.

Do có một vài điều kiện phức tạp ảnh hưởng đến bài toán và để xây dựng mô hình được thuận tiện hơn, báo cáo này đã đơn giản hóa một số giả thuyết sau:

- [1] Công ty A luôn có khả năng thanh toán khoản vay khi đến hạn.
- [2] Tại bất kỳ một thời điểm nào trong lúc hợp đồng còn hiệu lực, công ty A luôn có thể rút khoản vay có chu kỳ đáo hạn ít nhất 1 tháng và tối đa 14 tháng.
- [3] Công ty A chỉ rút tiền và trả tiền cuối tháng.
- [4] Lãi suất thị trường sẽ không biến động theo thời gian và mãi được duy trì ở 2.9%.
- [5] Thời gian đáo hạn trung bình là 3 tháng.
- [6] Số năm mô phỏng là 10 năm.
- [7] Số tiền giới hạn tín dụng là 1000 tỷ đồng.
- [8] Trung bình mỗi lần công ty A lần rút 100 tỷ đồng.

Để mô hình hóa hành vi rút tiền từ tài khoản hợp đồng tín dụng xoay vòng, ta xây dựng thuật toán để đồ thị hóa dữ liệu số tiền rút, thời hạn đáo hạn và số dư còn lại trong mỗi thời điểm. Đầu tiên ta sẽ sử dụng bộ dữ liệu số tiền rút, thời gian đáo hạn (1) có trong appendix, trong đó giá trị trung bình mỗi lần rút tiền là 100 (tỷ đồng), thời gian đáo hạn trung bình là 3 (tháng), lãi suất thị trường trung bình là 2.9% và dữ liệu bao gồm tổng cộng 120 tháng từ tháng 1 năm 2014 đến tháng 12 năm 2023 biết rằng trong bộ dữ liệu này, các lần rút được thực hiện ở cuối tháng. Ngoài ra, vì mô hình này xét cả trường hợp công ty A có thể vay nhiều lần với nhiều chu kỳ đáo hạn khác nhau miễn rằng đảm bảo được số dư trong tài khoản hợp đồng tín dụng xoay vòng không âm (ban đầu tài khoản này có 1000 tỷ). Từ đó ta xây dựng hàm tính toán số dư như sau:

$$B_t = B_{t-1} - W_t$$

Trong đó  $B_t$  là số dư tại tháng  $t$ ;  $W_t$  là số tiền rút tại tháng  $t$ .

Tiếp theo ta định nghĩa hàm đệ quy  $H(t)$  để cập nhật số dư như sau:

$$H(t) = \{(A_{t-i}, D_{t-i} - 1) | (A_{t-i}, D_{t-i}) \in H(t-1), D_{t-i} > 1\}$$

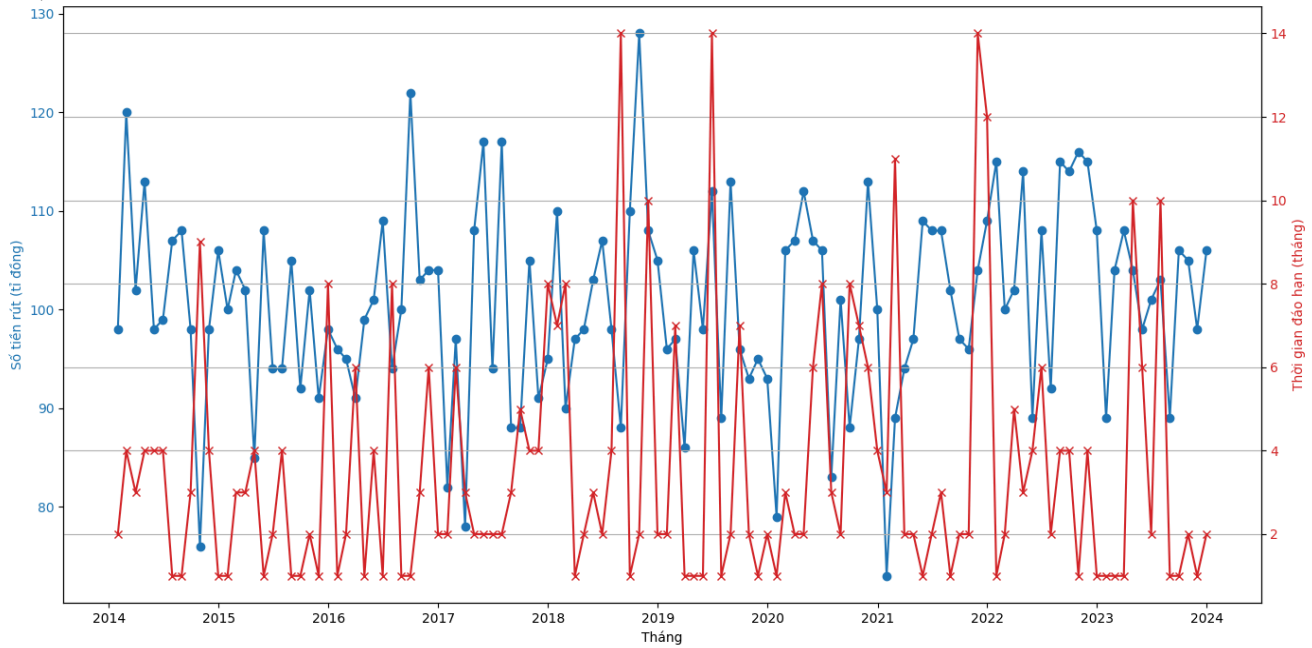
Trong đó, với mỗi tháng:  $A_{t-i}$  là số tiền có sẵn từ tháng trước đó;  $D_{t-i}$  là khoảng thời gian còn lại cho đến khi số tiền được trả lại vào số dư.

Khi  $D_{t-i} = 1$ , ta cộng các số dư lại:  $B_t = B_t + A_{t-i}$ .

Ngoài ra, ta tính lãi suất hằng tháng từ lãi suất hằng năm bằng công thức:  $I_t = (1 + I_n)^{\frac{1}{12}} - 1$  với  $I_t$  là lãi tháng và  $I_n$  là lãi năm.

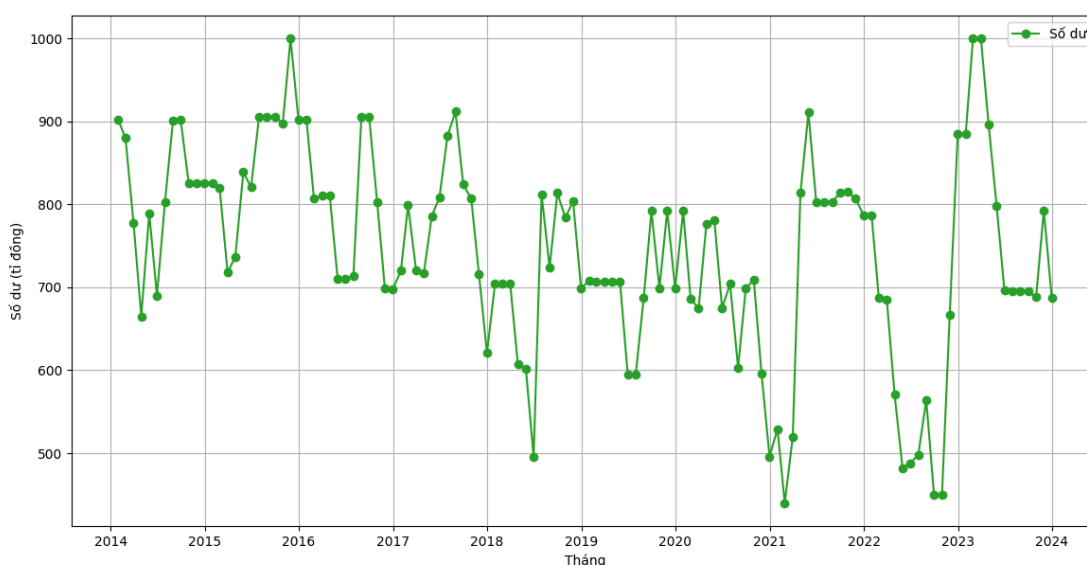
Cuối cùng, tổng lãi của công ty B được tính bằng công thức  $T_t = T_{t-1} + B_t \times I_t$ .

Để hoàn thành mô hình này, đầu tiên, chúng tôi đã triển khai thuật toán thành code Python (có thể xem ở phần Appendix). Sau đó, chúng tôi đã sử dụng thư viện matplotlib với Python 3.11 để đồ thị hóa dữ liệu, từ đó thu được các đồ thị sau:



Hình 2. Đồ thị tương quan giữa số tiền rút và thời gian đáo hạn trong 120 tháng





Hình 3. Số dư còn lại trong các khoảng thời gian

Mô hình này thể hiện số dư trong khoản vay tín dụng xoay vòng giúp công ty A có thể quản lý khoản vay dễ dàng và có chiến lược vay hợp lý nhất.

Từ hàm  $T_t$ , ta lấy  $t = 120$  là tháng cuối cùng trong hợp đồng 10 năm, ta tính được tổng lãi suất (giá trị hợp đồng) là khoảng 217,48 tỷ đồng. Mô hình này cũng có thể mô phỏng, đánh giá được hành vi của các công ty khác tùy vào thời gian, số tiền rút và dữ liệu đầu vào cho mô hình.

#### 4. Mô hình mô phỏng sự biến động của lãi suất và hành vi khách hàng

Data mining quá trình tự động lấy các mẫu trong các tập dữ liệu lớn, bao gồm phương pháp mô hình hóa dự đoán, trong đó ta có thể sử dụng các phương pháp regression. Báo cáo này sử dụng phương pháp phân chia đệ quy (Recursive Partitioning - RPART) để xây dựng mô hình hồi quy. Tiêu chí phân chia là SST – (SSL + SSR), trong đó SST là tổng bình phương sai số của nút, và SSR và SSL lần lượt là tổng bình phương sai số của nút con bên phải và bên trái. Nhận thấy rằng cây quyết định hồi quy (Decision Tree Regressor - DTR) là một phương pháp mạnh và đáng tin cậy trong các trường hợp như thiếu giá trị, giá trị phi tuyến hoặc chênh lệch bậc giữa các biến. Trong hợp đồng tín dụng xoay vòng (RCF), ta cần xác định yếu tố hoặc biến nào ảnh hưởng đến số tiền được rút trong hợp đồng (giá trị trung bình và phương sai hay độ lệch chuẩn của của khoản vay). Đầu tiên, ta xây dựng mô hình dự đoán dựa trên Decision Tree Regressor bằng cách thiết lập các thông số Entropy (mức độ ngẫu nhiên của tập dữ liệu), Information gain, chỉ số Gini, Gain Ratio, Reduction in Variance, Chi-Square.

Ta biểu diễn hàm đo mức độ ngẫu nhiên bằng hàm toán học sau:

$$E(S) = \sum_{i=1}^c -p_i \times \log_2(p_i)$$

Trong đó, Trong đó S là giá trị ban đầu,  $p_i$  là xác suất của sự kiện i (giá trị tăng hoặc giảm) của giá trị S. Tuy nhiên trong trường hợp này, ta còn cần xem xét tính ngẫu nhiên của tỷ lệ lãi suất trong từng thời điểm, vì thế ta sử dụng hàm biểu diễn hàm cho mức độ ngẫu nhiên của nhiều yếu tố:

$$E(T, X) = \sum_{c \in X} P(c) \times E(c)$$

Trong đó X là mức độ ngẫu nhiên của lãi suất thị trường.



Information gain đo mức độ tốt mà mức độ ngẫu nhiên của tập dữ liệu phân chia huấn luyện theo mục tiêu. Nói cách khác, information gain là sự giảm entropy. Hàm toán học Information được thể hiện như sau:

$$Information\ Gain(T, X) = Entropy(T) - Entropy(T, X)$$

Ta có thể đơn giản hóa hàm này thành:

$$Information\ Gain = Entropy(before) - \sum_{j=1}^k \times Entropy(j, after)$$

Chỉ số Gini nói cách khác là một hàm chi phí (cost function) sử dụng để đánh giá các phần nhỏ trong tập dữ liệu và được biểu diễn như sau:

$$Gini = 1 - \sum_{i=1}^C \times (p_i)^2$$

Gain Ratio là một hàm điều chỉnh Information Gain giúp tính toán những giá trị đặc biệt (phân tán hơn) trong tập dữ liệu và được biểu diễn bằng hàm:

$$Gain\ Ratio = \frac{Information\ Gain}{SplitInfo} = \frac{Entropy(before) - \sum_{j=1}^K Entropy(j, after)}{\sum_{j=1}^K w_j \times \log_2 w_j}$$

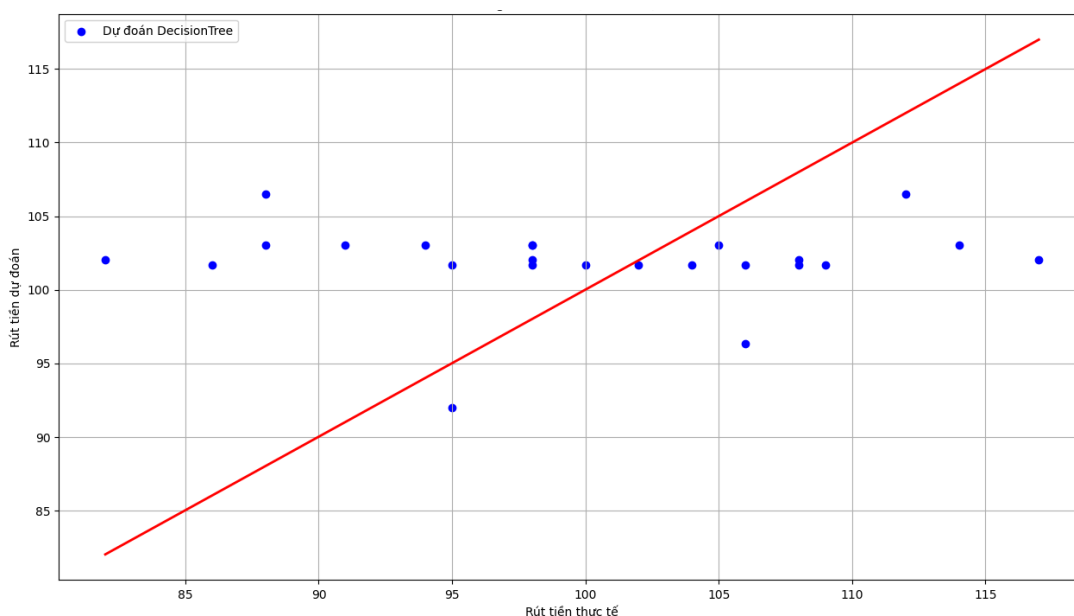
Reduction in variance là một thuật toán được sử dụng cho các biến xuất hiện liên tục trong hồi quy và được tính bằng hàm:

$$Variance = \frac{\sum (X - \bar{X})^2}{n}$$

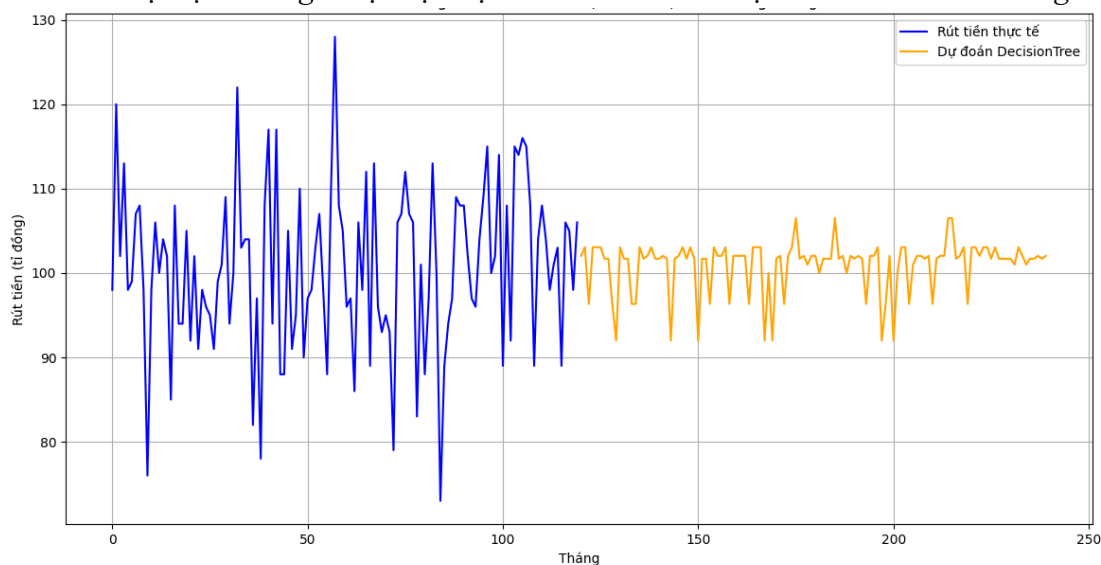
Chi-square là một phương pháp phân loại cây cổ giúp tìm ra sự khác biệt giữa các giá trị trong tập:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Cuối cùng, chúng tôi đã sử dụng thư viện matplotlib với Python 3.11 để đồ thị hóa dữ liệu và triển khai viết code cho thuật toán trên, từ đó thu được các đồ thị so sánh dữ liệu được dự đoán so với dữ liệu thực tế như sau:



Hình 4. Giá trị thực tế và giá trị được dự đoán bởi mô hình dựa trên *DecisionTreeRegressor*



Hình 5. Giá trị được dự đoán bởi mô hình dựa trên *DecisionTreeRegressor* trong 10 năm tới

Dựa vào dữ liệu đã có từ năm 2014 đến 2023 (120 tháng), chúng tôi đã áp dụng mô hình học máy Machine Learning cho thuật toán đã xây dựng trên tiến hành dự đoán hành vi của khách hàng dựa trên dữ liệu học được.

Nhận thấy rằng, sai số bình phương trung bình được tính bằng công thức sau của mô hình này còn lớn ( $\approx 91.90$ ):

$$\text{Mean squared error} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Trong đó,  $n$  là số thành phần có trong tập dữ liệu,  $y_i$  là giá trị thật,  $\hat{y}_i$  là giá trị dự đoán.

Chúng tôi đã thay đổi mô hình bằng cách cải tiến thuật toán có khả năng xử lý tập dữ liệu lớn hơn. Mô hình này đơn giản là một cải tiến của mô hình sử dụng thuật toán Decision Tree Regressor nhưng dữ liệu được chia nhỏ hơn và được xử lý bởi nhiều Decision Tree.

Tuy mô hình này sẽ cần nhiều tài nguyên tính toán hơn do chúng tôi đã sử dụng nhiều Decision Tree hơn, nhưng việc sử dụng thư viện Random Forest Regressor, mô hình mới có thể chia nhỏ dữ liệu và xử lý chúng song song, giúp tăng tốc độ huấn luyện và dự đoán. Ngoài ra, mô hình này cũng làm giảm hiện tượng overfit do mô hình này đã chia nhỏ tập dữ liệu để sử dụng phương pháp kết hợp nhiều decision tree, lấy mẫu ngẫu nhiên và giới hạn độ phức tạp của decision tree. Cuối cùng, mô hình mới này cũng cho ra độ chính xác cao hơn do có nhiều decision tree giúp giảm bias và phương sai trong dự đoán.

Do đó chúng tôi đề xuất thêm một mô hình dự đoán có tính chính xác cao hơn, trong trường hợp dữ liệu hành vi người dùng có tính phân bố lớn, được xây dựng như sau:

Đầu tiên ta tạo B tập dữ liệu con từ tập dữ liệu ban đầu  $D = \{(x_i, y_i)\}_{i=1}^N$ :

$$D^b = \{(x_i^b, y_i^b)\}_{i=1}^N$$

Trong đó  $b \in \{1, 2, \dots, B\}$

Tiếp theo, Với mỗi mẫu dữ liệu  $D^b$ , ta huấn luyện một cây quyết định  $T^b$ . Tại mỗi nút phân chia trong cây, chọn một tập con ngẫu nhiên các biến đầu vào  $M \subseteq \{1, 2, \dots, p\}$  với  $|M| = m$ , trong đó  $p$  là số lượng biến đầu vào ban đầu và  $m$  là số biến đầu vào được chọn ngẫu nhiên. Tại mỗi nút  $j$  trong cây, chọn biến  $k \in M$  và biến  $t$  để tối ưu hóa tiêu chí phân chia bằng hàm sau:

$$\min_{k,t} \left( \sum_{i \in \text{left}(j)} (y_i - \bar{y}_{\text{left}(j)})^2 + \sum_{i \in \text{right}(j)} (y_i - \bar{y}_{\text{right}(j)})^2 \right)$$

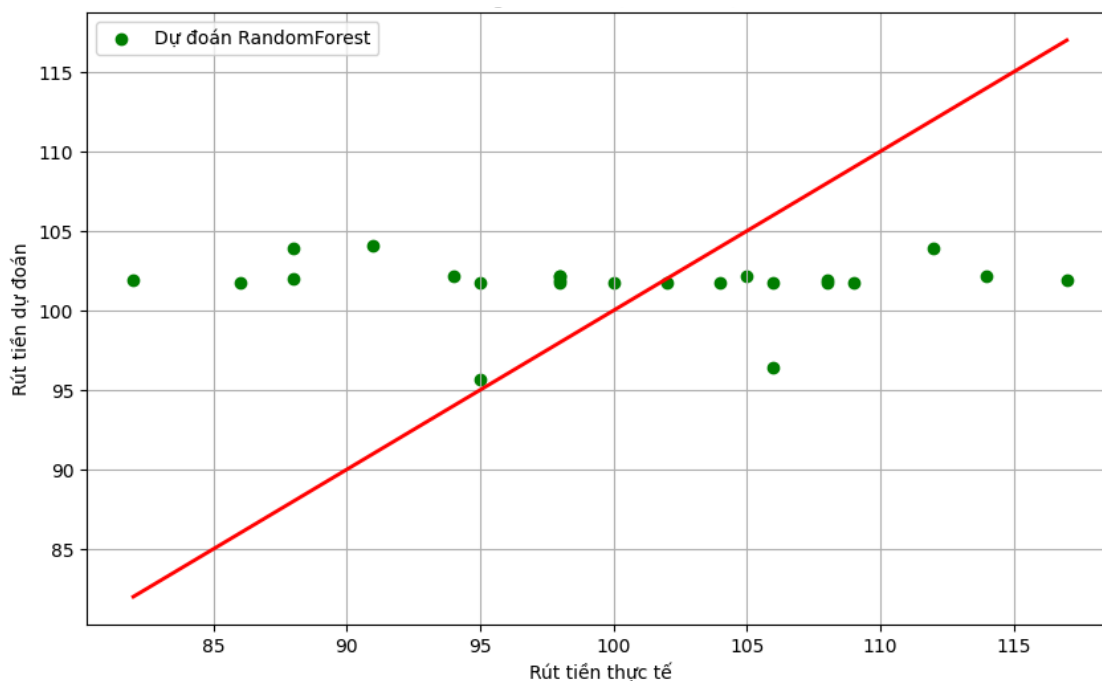
Với  $\text{left}(j)$  và  $\text{right}(j)$  là các tập hợp mẫu ở các nhánh trái và phải sau khi phân chia tại nút  $j$ , và  $\bar{y}_{\text{left}(j)}$ ,  $\bar{y}_{\text{right}(j)}$  là trung bình của các giá trị trong các tập đó. Sau khi huấn luyện xong, mỗi cây quyết định  $T^b$  sẽ dự đoán giá trị  $\hat{y}_b$  cho một mẫu  $x$  dựa trên dữ liệu đã học được:

$$T_b(x) = \hat{y}_b$$

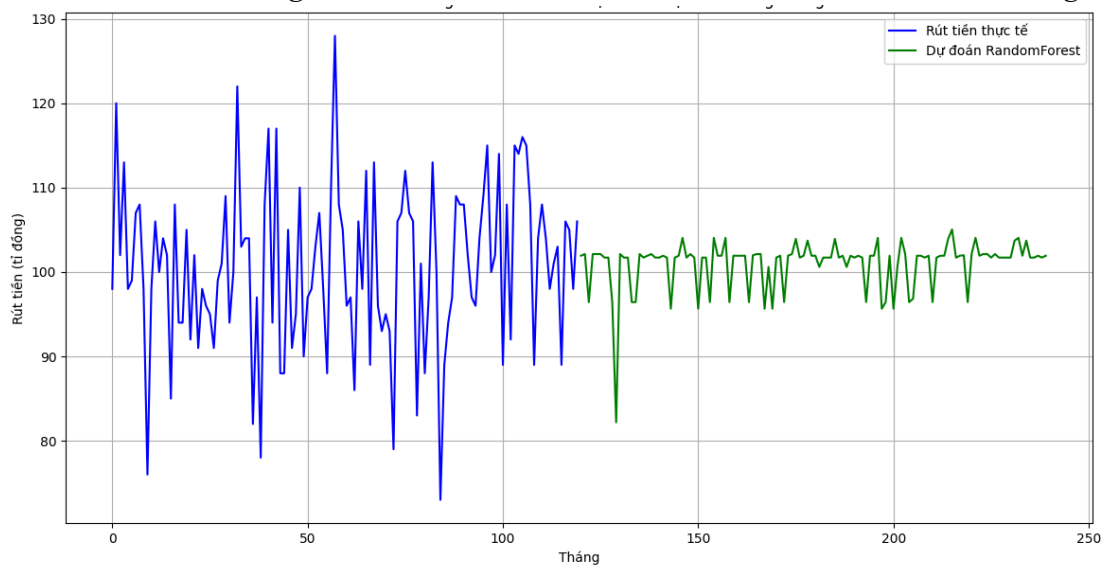
Cuối cùng, ta đưa ra kết quả dự đoán bằng cách lấy giá trị trung bình của tất cả các dự đoán dựa trên các cây:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B \hat{y}_b$$

Cuối cùng, chúng tôi đã sử dụng thư viện matplotlib với Python 3.11 để đồ thị hóa dữ liệu và triển khai viết code cho thuật toán trên, từ đó thu được các đồ thị so sánh dữ liệu được dự đoán bằng mô hình mới so với dữ liệu thực tế như sau:



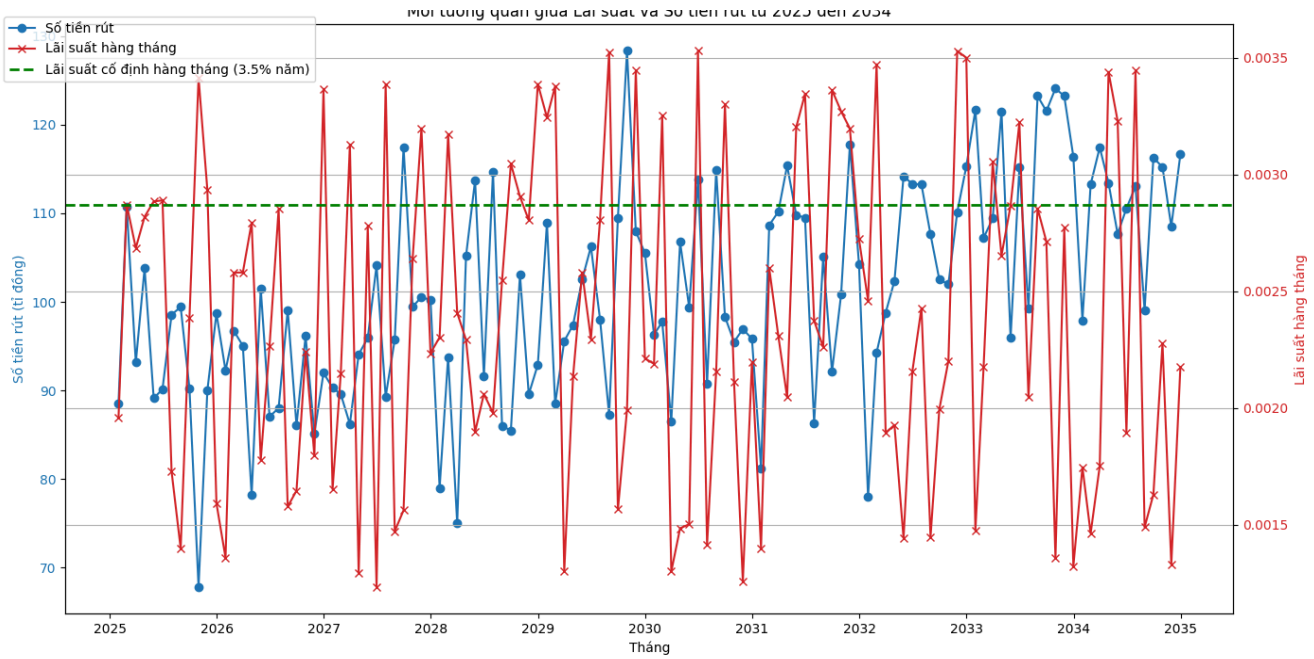
Hình 6. Giá trị thực tế và giá trị được dự đoán bởi mô hình dựa trên RandomForestRegressor



Hình 7. Giá trị được dự đoán bởi mô hình dựa trên RandomForestRegressor trong 10 năm tới

Nhận thấy rằng, sai số bình phương trung bình của mô hình này đã giảm còn 88,66. Từ đó, ta đánh giá được độ hiệu quả của mô hình sau cải tiến đã tăng lên và đưa ra kết quả chính xác hơn khi so sánh với dữ liệu thật, cũng như có thể đưa ra những dự báo hành vi khách hàng trong tương lai một cách đáng tin cậy hơn.

Tổng kết lại, ta có thể áp dụng mô hình đã cải tiến sử dụng thư viện Random Forest Regressor kết hợp với hàm  $H(t)$  tính toán giá trị hợp đồng bằng cách định nghĩa tổng lãi cho ngân hàng B dựa trên lãi suất hợp đồng. Điều này được mô phỏng trong đồ thị sau:



Hình 8. Mối tương quan giữa lãi suất và số tiền rút được đoán từ năm 2025 đến 2035 dựa trên dữ liệu đã có

Trong đó chúng tôi giả sử hợp đồng được ký với lãi suất cố định hằng năm là 2.9%, sau đó ta tính được lãi tháng bằng công thức  $I_t = (1 + I_n)^{\frac{1}{12}} - 1$  với  $I_t$  là lãi tháng và  $I_n$  là lãi năm. Sau đó ta dùng hàm  $H(t)$  đã định nghĩa ở phần 3 và tính được tổng lãi là khoảng 217,48 tỷ đồng.

Nhìn vào biểu đồ ra có nhận xét sau: khi mức lãi suất thị trường hàng năm (3.5%) lớn hơn mức lãi suất hợp đồng kết hợp với việc ngân hàng rút số tiền lớn sẽ gây bất lợi cho ngân hàng trong việc huy động vốn và gây thiệt hại cho ngân hàng B. Ngược lại, nếu công ty A rút tiền trong giai đoạn lãi suất thị trường thấp sẽ giúp ngân hàng B có lợi hơn và thu được lợi nhuận cao. Từ biểu đồ trên, công ty A rút tiền vào các thời điểm mà mức lãi suất thị trường dưới mức lãi suất hợp đồng chiếm phần lớn hơn so với việc công ty A rút tiền vào thời điểm mức lãi suất thị trường cao hơn mức lãi suất hợp đồng. Tuy nhiên, việc công ty A rút tiền vào các thời điểm mức lãi suất cao hơn thị trường cũng là đáng kể, do đó việc mô hình hóa hành vi của công ty A như trên có thể giúp ngân hàng B đưa ra được mức lãi suất phù hợp.

## IV. ĐÁNH GIÁ MÔ HÌNH

### 1. Nhận xét chung về bài toán

Bài toán mà mô hình này đang giải quyết là dự đoán hành vi rút tiền của công ty A từ tài khoản tín dụng xoay vòng, nhằm định giá hợp đồng tín dụng này một cách chính xác và hợp lý. Mục tiêu chính là xác định quy luật rút tiền của công ty A, bao gồm các yếu tố như thời điểm rút tiền, thời gian đáo hạn mỗi lần rút, và số tiền rút mỗi lần.

Bài toán này rất quan trọng trong bối cảnh thực tế vì nó giúp ngân hàng B tối ưu hóa quản lý dòng tiền và giảm thiểu rủi ro trong việc cung cấp tín dụng. Các dự báo chính xác về hành vi rút tiền của khách hàng giúp ngân hàng đưa ra các quyết định chiến lược về lãi suất và kỳ hạn cho vay, đồng thời đảm bảo lợi nhuận và khả năng cạnh tranh trên thị trường tài chính. Do đó, mô hình này có thể mở rộng ứng dụng trong nhiều lĩnh vực khác nhau, đặc biệt là trong quản lý tín dụng và tài chính doanh nghiệp.

## 2. Mô hình thứ nhất

Mô hình thứ nhất được xây dựng dựa trên phương pháp Decision Tree Regressor (DTR) để mô phỏng sự biến động của lãi suất và hành vi khách hàng. Phương pháp này có khả năng xử lý các dữ liệu có giá trị bị thiếu hoặc phi tuyến tính. Decision Tree Regressor giúp xác định rõ các yếu tố ảnh hưởng đến số tiền rút trong hợp đồng tín dụng xoay vòng, từ đó dự đoán chính xác hơn về hành vi rút tiền của khách hàng. Mô hình này cũng có khả năng phân chia dữ liệu dựa trên các thông số như Entropy, Information Gain, và Chi-Square, giúp đưa ra những dự báo về tình hình tài chính của công ty A.

## 3. Mô hình thứ hai

Mô hình thứ hai được xây dựng dựa trên Random Forest Regressor để dự đoán hành vi rút tiền của công ty A. Mô hình này đã cải tiến được độ chính xác đáng kể so với các phương pháp truyền thống. Cụ thể, sai số bình phương trung bình của mô hình này giảm xuống còn 88,66, cho thấy sự hiệu quả trong việc dự đoán hành vi khách hàng và đưa ra những dự báo tin cậy hơn về tương lai. Việc sử dụng Random Forest Regressor giúp mô hình hóa tốt các yếu tố phi tuyến tính và biến động của thị trường hơn nhờ sử dụng nhiều node của nhiều Decision Tree hơn, từ đó dự đoán, phân tích chính xác hơn về lợi ích tài chính của ngân hàng.

## 4. So sánh 2 mô hình

So sánh giữa hai mô hình Random Forest Regressor và Decision Tree Regressor, mỗi mô hình đều có những đặc điểm riêng biệt và ưu nhược điểm cụ thể. Random Forest Regressor vượt trội hơn trong việc giảm thiểu sai số dự báo nhờ khả năng xử lý tốt các yếu tố phi tuyến và biến động dữ liệu. Mô hình này cũng tận dụng hiệu quả dữ liệu lịch sử để đưa ra các dự báo tin cậy. Tuy nhiên, nó đòi hỏi tài nguyên tính toán cao hơn và có thể trở nên phức tạp trong quá trình triển khai. Ngược lại, Decision Tree Regressor dễ hiểu và triển khai hơn, đồng thời cũng đảm bảo xử lý tốt các dữ liệu thiếu và phi tuyến tính ở một mức độ hạn chế. Tuy nhiên, mô hình này dễ bị overfitting nếu không được tối ưu hóa cẩn thận và có thể không đạt được độ chính xác cao như Random Forest Regressor trong các tình huống phức tạp. Mặc dù Random Forest Regressor nổi trội hơn so với Decision Tree Regressor, nó vẫn chưa tối ưu hoá sai số dự dẫn đến hiện tượng sai lệch cao hơn và các phán đoán không chính xác.

## V. TÍNH ỨNG DỤNG CỦA MÔ HÌNH BÀI TOÁN

Các mô hình được xây dựng trong báo cáo có tính ứng dụng cao trong lĩnh vực tài chính, đặc biệt là trong việc quản lý tín dụng xoay vòng. Cả hai mô hình giúp ngân hàng B có thể dự báo chính xác hành vi rút tiền của khách hàng, từ đó tối ưu hóa các quyết định về lãi suất hợp đồng và quản lý rủi ro tài chính hiệu quả hơn. Điều này không chỉ giúp ngân hàng đảm bảo lợi nhuận mà còn nâng cao khả năng cạnh tranh trên thị trường. Ngoài ra, mô hình còn có thể mở rộng để áp dụng cho nhiều công ty và ngành nghề khác nhau, tùy thuộc vào dữ liệu đầu vào, góp phần nâng cao hiệu quả quản lý tài chính và dự báo trong các tổ chức tài chính khác.

## VI. TRÍCH DẪN - TÀI LIỆU THAM KHẢO

- [1] Các tài liệu của cuộc thi VM2C
  - [2] Financial data (2024). Dữ liệu tỷ giá - lãi suất Việt Nam. Retrieved from <https://finance.vietstock.vn/du-lieu-vi-mo/53-64/ty-gia-lai-suat.htm>
  - [3] Mirkov, R., Thomae, H., Feist, M., Maul, T., Gillespie, G., & Lie, B. (2014). Modeling and Forecasting Customer Behavior for Revolving Credit Facilities.
  - [4] Sanchez-Barrios, L. J., Andreeva, G., & Ansell, J. (2016). Time-to-profit scorecards for revolving credit. *European Journal of Operational Research*, 249(2), 397-406.
  - [5] Paxton, J., Graham, D., & Thraen, C. (2000). Modeling group loan repayment behavior: New insights from Burkina Faso. *Economic Development and cultural change*, 48(3), 639-655.
  - [6] Alimov, D. A., Obrosova, N. K., & Shananin, A. A. (2020). Mathematical Model of Enterprise with Revolving Funds Deficit: Analysis of Demand Shocks 2020. *Lobachevskii Journal of Mathematics*, 41(12), 2628–2637. <https://doi.org/10.1134/S1995080220120045>
  - [7] WallStreetPrep (2023). Revolver debt - fundamentals of financial modeling. Retrieved from <https://www.wallstreetprep.com/knowledge/modeling-revolving-credit-line-excel-free-template/>
  - [8] Eloquens, WallStreetPrep (2018). Revolving Credit Line Excel Model. Retrieved from <https://www.eloquens.com/tool/eRrFXn/finance/all/revolver-revolving-credit-line-excel-model-exercis-e-template>
-



## VII. PHỤ LỤC

[1] Source code chương trình chạy của nhóm chúng tôi có thể được tìm thấy ở đường link này:

[https://github.com/dngvmnh/VM2C\\_202400000471](https://github.com/dngvmnh/VM2C_202400000471)

[2] Bảng dữ liệu đã dùng trong bài báo cáo:

Thang	Tiền rút	Thời gian	Số dư	Lai thang	Lai nam
31/01/2025	98	2	902	0.002402	0.029203
28/02/2025	120	4	880	0.002451	0.029808
31/03/2025	102	3	778	0.002447	0.029768
30/04/2025	113	4	665	0.002503	0.030448
31/05/2025	98	4	789	0.002392	0.029084
30/06/2025	99	4	690	0.002463	0.029962
31/07/2025	107	1	803	0.002438	0.029647
31/08/2025	108	1	901	0.002448	0.02978
30/09/2025	98	3	902	0.002336	0.028401
31/10/2025	76	9	826	0.002401	0.029192
30/11/2025	98	4	826	0.002458	0.029894
31/12/2025	106	1	826	0.002396	0.029131
31/01/2026	100	1	826	0.002469	0.03003
28/02/2026	104	3	820	0.002386	0.029012
31/03/2026	102	3	718	0.002327	0.02828
30/04/2026	85	4	737	0.002294	0.027873
31/05/2026	108	1	839	0.002328	0.028297
30/06/2026	94	2	821	0.002413	0.029342
31/07/2026	94	4	906	0.002376	0.028885
31/08/2026	105	1	906	0.002442	0.029698
30/09/2026	92	1	906	0.002356	0.028643

31/10/2026	102	2	898	0.002449	0.029784
30/11/2026	91	1	1000	0.002343	0.028481
31/12/2026	98	8	902	0.002497	0.030382
31/01/2027	96	1	902	0.002338	0.028416
28/02/2027	95	2	807	0.002471	0.030055
31/03/2027	91	6	811	0.002369	0.028796
30/04/2027	99	1	811	0.002325	0.02826
31/05/2027	101	4	710	0.002493	0.030328
30/06/2027	109	1	710	0.00236	0.028686
31/07/2027	94	8	714	0.002437	0.029637
31/08/2027	100	1	906	0.002306	0.028031
30/09/2027	122	1	906	0.002494	0.030346
31/10/2027	103	3	803	0.0023	0.027952
30/11/2027	104	6	699	0.002473	0.030089
31/12/2027	104	2	698	0.002271	0.027593
31/01/2028	82	2	720	0.002342	0.028469
29/02/2028	97	6	799	0.002379	0.02893
31/03/2028	78	3	721	0.002436	0.029628
30/04/2028	108	2	717	0.002432	0.029574
31/05/2028	117	2	786	0.002333	0.028354
30/06/2028	94	2	809	0.002285	0.027767
31/07/2028	117	2	883	0.00229	0.027832
31/08/2028	88	3	912	0.002267	0.02755
30/09/2028	88	5	824	0.002447	0.029767
31/10/2028	105	4	807	0.002299	0.027943

30/11/2028	91	4	716	0.002459	0.029912
31/12/2028	95	8	621	0.002333	0.028353
31/01/2029	110	7	704	0.002309	0.028059
28/02/2029	90	8	705	0.002377	0.028895
31/03/2029	97	1	705	0.002441	0.02969
30/04/2029	98	2	607	0.002467	0.030006
31/05/2029	103	3	602	0.00239	0.029061
30/06/2029	107	2	495	0.002336	0.028391
31/07/2029	98	4	812	0.002367	0.02878
31/08/2029	88	14	724	0.002492	0.030319
30/09/2029	110	1	814	0.002456	0.029874
31/10/2029	128	2	784	0.002357	0.028659
30/11/2029	108	10	804	0.002431	0.029569
31/12/2029	105	2	699	0.002368	0.028784
31/01/2030	96	2	708	0.002451	0.02981
28/02/2030	97	7	707	0.002335	0.028378
31/03/2030	86	1	707	0.002365	0.028753
30/04/2030	106	1	707	0.002413	0.029347
31/05/2030	98	1	707	0.002304	0.027997
30/06/2030	112	14	595	0.002439	0.029664
31/07/2030	89	1	595	0.002424	0.029483
31/08/2030	113	2	687	0.0025	0.03041
30/09/2030	96	7	792	0.002437	0.029644
31/10/2030	93	2	699	0.002414	0.02936
30/11/2030	95	1	792	0.002406	0.029254

31/12/2030	93	2	699	0.002421	0.029439
31/01/2031	79	1	792	0.002493	0.03033
28/02/2031	106	3	686	0.002293	0.027869
31/03/2031	107	2	675	0.002476	0.030125
30/04/2031	112	2	776	0.002343	0.028483
31/05/2031	107	6	781	0.002268	0.027563
30/06/2031	106	8	675	0.002323	0.028231
31/07/2031	83	3	704	0.002391	0.029074
31/08/2031	101	2	603	0.002398	0.029155
30/09/2031	88	8	699	0.002396	0.029137
31/10/2031	97	7	709	0.002406	0.029263
30/11/2031	113	6	596	0.002275	0.027643
31/12/2031	100	4	496	0.002503	0.030448
31/01/2032	73	3	529	0.002368	0.028787
29/02/2032	89	11	440	0.002305	0.028013
31/03/2032	94	2	519	0.00232	0.028196
30/04/2032	97	2	814	0.002437	0.029643
31/05/2032	109	1	911	0.002346	0.028516
30/06/2032	108	2	803	0.0024	0.029182
31/07/2032	108	3	803	0.002492	0.030314
31/08/2032	102	1	803	0.002377	0.028899
30/09/2032	97	2	814	0.002399	0.029175
31/10/2032	96	2	815	0.00234	0.02845
30/11/2032	104	14	807	0.002439	0.029661
31/12/2032	109	12	787	0.002437	0.02964

31/01/2033	115	1	787	0.00236	0.028692
28/02/2033	100	2	687	0.002356	0.028643
31/03/2033	102	5	685	0.002426	0.029507
30/04/2033	114	3	571	0.002426	0.029506
31/05/2033	89	4	482	0.002467	0.030008
30/06/2033	108	6	488	0.002496	0.030362
31/07/2033	92	2	498	0.002337	0.028412
31/08/2033	115	4	564	0.002348	0.028537
30/09/2033	114	4	450	0.002353	0.028601
31/10/2033	116	1	450	0.002403	0.029214
30/11/2033	115	4	667	0.002335	0.028385
31/12/2033	108	1	885	0.002408	0.029277
31/01/2034	89	1	885	0.002393	0.029091
28/02/2034	104	1	1000	0.002313	0.028111
31/03/2034	108	1	1000	0.00245	0.029803
30/04/2034	104	10	896	0.002448	0.029773
31/05/2034	98	6	798	0.002289	0.027818
30/06/2034	101	2	697	0.002406	0.029261
31/07/2034	103	10	695	0.002418	0.0294
31/08/2034	89	1	695	0.002291	0.027841
30/09/2034	106	1	695	0.002319	0.028183
31/10/2034	105	2	688	0.002439	0.02967
30/11/2034	98	1	793	0.002322	0.028219
31/12/2034	106	2	687	0.002312	0.028104

**VIII. LỜI CẢM ƠN**

Đầu tiên, chúng tôi xin chân thành gửi lời cảm ơn đến Ban tổ chức VM2C vì đã mang đến cho học sinh một cuộc thi đầy ý nghĩa, đây là một cơ hội rất lớn cho học sinh học tập lẫn trải nghiệm thực tế và có khoảng thời gian học nhóm, thảo luận sôi nổi. Sau khi tham gia cuộc thi này, chúng tôi không chỉ được tập huấn với những kiến thức mới lạ, những mẹo bổ ích qua những buổi tập huấn bởi ban chuyên môn mà còn hiểu thêm về những ứng dụng của bộ môn toán mô hình và những cơ hội nghề nghiệp lớn dành cho chúng tôi sau này. Không chỉ có thế, chương trình đã là một bước đệm giúp chúng tôi đạt được ước mơ mà chúng tôi sẽ không thể nào quên.

Tiếp đến, chúng tôi muốn lời cảm ơn chân thành nhất đến ban cố vấn của cuộc thi VM2C, các thầy cô ở viện toán cao cấp VIASM đã giúp đỡ nhóm trong suốt quá trình tập huấn và tham gia chương trình, đặc biệt là giáo viên của trường đã giúp chúng tôi biết đến cuộc thi và chỉ dạy từ đầu. Mọi người chính là nguồn động lực của chúng tôi hoàn thành bản báo cáo này. Cuối cùng, nhóm xin cảm ơn trường Đại học Khoa học Tự nhiên cùng với các đơn vị tài trợ khác đã tạo điều kiện để cuộc thi có thể diễn ra thành công tốt đẹp.

Chúng tôi trân trọng mọi sự đóng góp và hướng dẫn của tất cả những người đã cùng đồng hành cùng chúng tôi để có những kỉ niệm đáng quý tại cuộc thi. Chúng tôi tin rằng, thông qua chương trình này, các em học sinh sẽ được trang bị những kỹ năng và kiến thức cần thiết để tự tin bước vào những thử thách mới trong tương lai. Một lần nữa, xin chân thành cảm ơn Ban Tổ chức vì những nỗ lực và tâm huyết đã dành cho chương trình Toán Mô Hình Việt Nam 2024.

---