

## APPLICATION OF THE R LANGUAGE IN DEVELOPING DIGITAL DOCUMENTATION FOR INFORMATION TECHNOLOGY STUDENTS

Doan Quang Duy<sup>1</sup>, Tran Quang Quy<sup>2\*</sup>, Le Anh Tu<sup>2</sup>, Phan Thi Cuc<sup>2</sup>  
Mong Quoc Tuan<sup>2</sup>, Dang Thuong Hoai Linh<sup>2</sup>, Le Hong Thu<sup>2</sup>

<sup>1</sup>Ho Chi Minh Communist Youth Union of Thai Nguyen province

<sup>2</sup>TNU - University of Information and Communication Technology

ARTICLE INFO	ABSTRACT
<b>Received:</b> 26/02/2024	Currently, there is a wide range of diverse and extensive learning materials available in the field of information technology. These include traditional textbooks, online resources, scientific articles, specialized journals, and instructional materials from technology service providers and manufacturers. However, the selection and evaluation of learning materials in the field of information technology pose a challenge due to the rapid development of technology and continuous changes in this field. Materials can quickly become outdated or no longer suitable. The purpose of the research in this article is to propose an approach to utilizing current open-source tools for building digital learning materials, enabling students to easily acquire knowledge during their study and research in the field of information technology. In this article, the authors present the application of the R and Markdown languages and the "reproducible research" method to refine a digital documentation process. The results of the research include digital learning materials that are currently being used by students majoring in Information Technology at the University of Information and Communication Technology - Thai Nguyen University.
<b>Revised:</b> 31/3/2024	
<b>Published:</b> 31/3/2024	

### KEYWORDS

Digital documentation  
Information Technology  
R language  
Markdown  
Reproducibility

## ỨNG DỤNG NGÔN NGỮ R XÂY DỰNG TÀI LIỆU SỐ DÀNH CHO SINH VIÊN CHUYÊN NGÀNH CÔNG NGHỆ THÔNG TIN

Đoàn Quang Duy<sup>1</sup>, Trần Quang Quý<sup>2\*</sup>, Lê Anh Tú<sup>2</sup>, Phan Thị Cúc<sup>2</sup>  
Mông Quốc Tuấn<sup>2</sup>, Đặng Thương Hoài Linh<sup>2</sup>, Lê Hồng Thu<sup>2</sup>

<sup>1</sup>Đoàn thanh niên cộng sản Hồ Chí Minh tỉnh Thái Nguyên

<sup>2</sup>Trường Đại học Công nghệ thông tin và Truyền thông - ĐH Thái Nguyên

THÔNG TIN BÀI BÁO	TÓM TẮT
<b>Ngày nhận bài:</b> 26/02/2024	Hiện nay, tài liệu học về công nghệ thông tin rất đa dạng và phong phú. Bao gồm trong đó từ sách giáo trình truyền thống cho đến tài liệu trực tuyến, các bài viết khoa học, tạp chí chuyên ngành và tài liệu hướng dẫn từ các nhà cung cấp dịch vụ và nhà sản xuất công nghệ. Tuy nhiên, việc lựa chọn và đánh giá tài liệu học trong lĩnh vực công nghệ thông tin cũng đặt ra một thách thức đó là sự phát triển nhanh chóng của công nghệ và sự thay đổi liên tục trong lĩnh vực này, tài liệu có thể nhanh chóng trở nên lỗi thời hoặc không còn phù hợp. Mục đích của nghiên cứu trong bài báo này nhằm đưa ra phương pháp tiếp cận các công cụ mã nguồn mở hiện nay trong việc xây dựng tài liệu số giúp sinh viên có thể dễ dàng tiếp thu kiến thức trong quá trình học và nghiên cứu chuyên ngành về công nghệ thông tin. Trong bài viết này, nhóm tác giả trình bày về ứng dụng của ngôn ngữ R và Markdown và phương pháp "nghiên cứu khả lặp" để hoàn thiện một quy trình xây dựng tài liệu số. Kết quả của nghiên cứu là các tài liệu số đã và đang được sử dụng cho sinh viên chuyên ngành Công nghệ thông tin tại Trường Đại học Công nghệ Thông tin và Truyền thông - Đại học Thái Nguyên.
<b>Ngày hoàn thiện:</b> 31/3/2024	
<b>Ngày đăng:</b> 31/3/2024	

### TỪ KHÓA

Tài liệu số  
Công nghệ thông tin  
Ngôn ngữ R  
Markdown  
Khả lặp

DOI: <https://doi.org/10.34238/tnu-jst.9781>

\* Corresponding author. Email: tqquy@ictu.edu.vn

## 1. Giới thiệu

Một trong những khó khăn chính khi học công nghệ thông tin là tốc độ phát triển nhanh chóng của lĩnh vực này bên cạnh sự gia tăng theo cấp số mũ của dữ liệu lớn [1]. Tốc độ phát triển này có thể dẫn đến sự lạc hậu và lỗi thời về công nghệ của bất kỳ nội dung hoặc kiến thức nào mà người học tìm hiểu được khi tiếp cận, yêu cầu đặt ra với người học về tính kiên trì và khả năng cập nhật công nghệ mới trong lĩnh vực công nghệ thông tin luôn cần thiết hơn bao giờ hết. Hơn nữa, công nghệ thông tin yêu cầu một sự tập trung cao đối với chi tiết và sự chính xác. Một sai sót nhỏ có thể dẫn đến hậu quả lớn trong việc phát triển phần mềm hoặc quản lý hệ thống [2]. Do đó, người học cần có khả năng kiên nhẫn và sự tỉ mỉ trong quá trình học tập và làm việc.

Lập trình là một kỹ năng quan trọng và không thể thiếu trong ngành công nghệ thông tin. Sinh viên thường gặp phải nhiều thách thức khi học và thực hành lập trình, một trong các lý do có thể kể đến như:

1. Khó khăn về cú pháp và ngôn ngữ lập trình: Mỗi ngôn ngữ lập trình đều có cú pháp và quy tắc riêng. Sinh viên mới học lập trình thường phải đối mặt với việc hiểu và sử dụng đúng cú pháp của ngôn ngữ. Sự thiếu thông tin hoặc sai sót nhỏ trong cú pháp có thể dẫn đến lỗi trong chương trình [3].

2. Quản lý thời gian: Lập trình yêu cầu sự kiên nhẫn và tập trung. Sinh viên thường phải dành nhiều thời gian để nghiên cứu và thực hành lập trình. Tuy nhiên, việc quản lý thời gian hiệu quả là một thách thức đối với sinh viên, đặc biệt khi họ phải xử lý nhiều bài tập và dự án cùng một lúc [4].

3. Cập nhật công nghệ mới: Lĩnh vực lập trình liên tục phát triển và cập nhật. Sinh viên cần theo dõi và nắm bắt những công nghệ mới, thay đổi trong ngành để không bị lạc hậu. Điều này đòi hỏi sinh viên phải có khả năng tự học và nghiên cứu liên tục để cập nhật kiến thức và kỹ năng [5].

Việc xây dựng một tài liệu số phù hợp với chuyên ngành công nghệ thông tin và đảm bảo tính đầy đủ, diễn tả lại được quy trình và các thao tác lập trình giúp sinh viên hiểu rõ và nắm bắt được toàn bộ quá trình của một bài toán lập trình luôn là những thách thức rất lớn đối với người học. Một khía cạnh khó khăn khác khi sử dụng tài liệu số là thiếu sự tương tác và phản hồi trực tiếp từ giảng viên. Sinh viên có thể gặp khó khăn trong việc giải đáp thắc mắc và vướng phải những nội dung không có người giải đáp. Với các nội dung như trên, việc xây dựng tài liệu số phù hợp bao gồm các thông tin cần thiết đối với người học, giúp người học tự có khả năng học tập, thực hành là điều vô cùng cần thiết. Xây dựng tài liệu số theo phương pháp nghiên cứu khả lập là một trong các phương pháp được đánh giá tốt và mang lại sự chủ động đối với sinh viên khi học tập, nghiên cứu [6] - [8]. Mục đích của nghiên cứu này là tìm hiểu về ứng dụng của ngôn ngữ phân tích thống kê số liệu R và ngôn ngữ Markdown trong tạo ra các dạng bài học có tính khả lập và có khả năng lưu trữ ở nhiều định dạng khác nhau.

## 2. Phương pháp nghiên cứu

### 2.1. Khái niệm về nghiên cứu khả lập

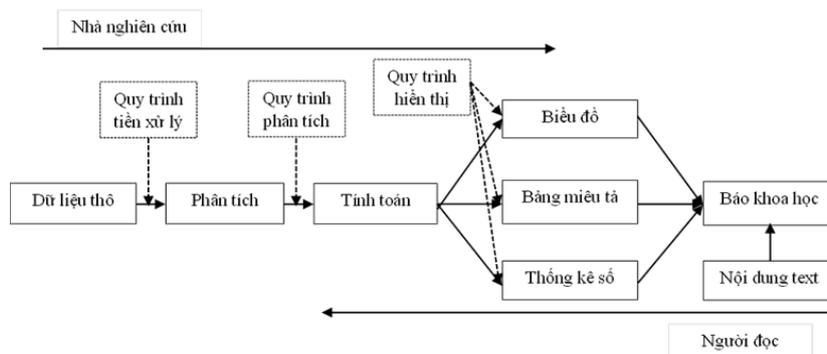
Phương pháp nghiên cứu theo hướng khả lập không chỉ có báo cáo kết quả nghiên cứu mà còn bao gồm đầy đủ các bước thực hiện, đồng thời công bố các thông tin cần thiết (dữ liệu, thuật toán sử dụng, ...) để người khác có thể tiến hành lại nghiên cứu nhằm xác nhận tính đúng đắn và mở rộng nghiên cứu dựa trên nền tảng nghiên cứu cũ. Trong trường hợp này, "khả lập" ("reproducible" hay "reproducibility") mang nghĩa có khả năng lặp lại, có khả năng thực hiện lại các nội dung đã được công bố [7], [8].

Theo mô hình nghiên cứu khả lập (Hình 1) do Roger D. Peng [9] đề xuất, người làm nghiên cứu sẽ định ra các bước làm theo một trình tự nhất định:

- Đặt ra câu hỏi cho bài toán nghiên cứu, tìm hiểu nguồn lấy dữ liệu và thu thập dữ liệu thô.
- Chuẩn bị dữ liệu, thao tác tiền xử lý và phân tích dữ liệu.
- Đưa ra các phương pháp tính toán thống kê cần thiết, đáp ứng câu hỏi của bài toán nghiên cứu, đưa ra các thư viện thao tác dữ liệu cần thiết và các bước chi tiết trong phân tích dữ liệu.
- Thực hiện các tính toán thống kê cần thiết và ước lượng các tham số.

- Hiện thị kết quả đạt được thông qua một số phương thức như: Dạng bảng, dạng biểu đồ trực quan, dạng báo cáo tóm tắt thống kê số tổng quan.

- Đưa kết quả hiển thị tới người đọc thông qua dạng báo cáo khoa học, dạng nội dung text. Người đọc khi tiếp cận quy trình phát triển nghiên cứu khả lập có khả năng tái tạo lại được các kết quả nghiên cứu trước đó, thực hiện các nội dung từ các quy trình nhỏ nhất tới các mô hình tính toán phức tạp.



**Hình 1.** Quy trình phát triển của nghiên cứu khả lập

Áp dụng nghiên cứu khả lập trong việc xây dựng tài liệu số là một hướng tiếp cận khoa học và hiệu quả để cải thiện quá trình và chất lượng của tài liệu. Phương pháp khả lập cho phép các nhà nghiên cứu và nhà phát triển tài liệu tiến hành lặp lại các giai đoạn quan trọng trong quy trình xây dựng, đánh giá và cải tiến nội dung, cấu trúc và hình thức truyền tải của tài liệu. Điều quan trọng khác mà nghiên cứu khả lập mang lại là khả năng thích nghi với sự tiến bộ của công nghệ. Trong lĩnh vực xây dựng tài liệu số, công nghệ không ngừng phát triển và mang đến những công cụ và phương pháp mới. Áp dụng nghiên cứu khả lập cho phép các nhà nghiên cứu, giảng viên thử nghiệm và đánh giá các công nghệ mới như công cụ tạo tài liệu, hình thức truyền tải và nền tảng trực tuyến. Điều này giúp tạo ra tài liệu số linh hoạt, dễ tiếp cận và phù hợp với sự phát triển của công nghệ [10]. Bên cạnh đó, giảng viên tạo học liệu số từ phương pháp khả lập có rất nhiều lợi ích mang lại:

1. Bằng cách áp dụng phương pháp nghiên cứu khả lập, giảng viên có thể lặp lại các giai đoạn trong quy trình xây dựng tài liệu, như thu thập thông tin, chỉnh sửa code trực tiếp và tái tạo văn bản,... Điều này giúp cải tiến quy trình để tăng tính hiệu quả và chất lượng của tài liệu.

2. Sử dụng phương pháp khả lập, giảng viên có thể tiến hành các phiên bản tài liệu khác nhau với nội dung và cấu trúc khác nhau. Qua quá trình đánh giá và cải tiến, giảng viên có thể tìm ra cách tốt nhất để tổ chức và trình bày thông tin trong tài liệu để đạt được sự rõ ràng và dễ tiếp cận cho sinh viên.

3. Áp dụng nghiên cứu khả lập cho xây dựng tài liệu số cho phép giảng viên thử nghiệm và đánh giá những công nghệ mới, như công cụ xây dựng tài liệu, hình thức truyền tải (ví dụ: ebook, ứng dụng di động) và nền tảng trực tuyến. Điều này giúp cập nhật và áp dụng những xu hướng công nghệ mới vào công việc xây dựng tài liệu.

## 2.2. Tìm hiểu về R và Markdown

Ngôn ngữ R là một ngôn ngữ lập trình và môi trường tính toán phổ biến trong lĩnh vực thống kê và khoa học dữ liệu. R được phát triển bởi các nhà khoa học và kỹ sư trong cộng đồng nghiên cứu, và nó đã trở thành một công cụ quan trọng cho việc phân tích số liệu, xử lý dữ liệu và biểu đồ hóa kết quả nghiên cứu. Một trong những đặc điểm nổi bật của ngôn ngữ R là tính linh hoạt và tích hợp sẵn của nó. R cung cấp một loạt các gói phần mềm và thư viện mạnh mẽ, giúp người dùng thực hiện các phương pháp thống kê phức tạp và mô hình hóa dữ liệu. Điều này cho phép các nhà nghiên cứu và nhà phân tích dữ liệu tùy chỉnh và mở rộng khả năng tính toán của môi trường R theo nhu cầu cụ thể của họ.

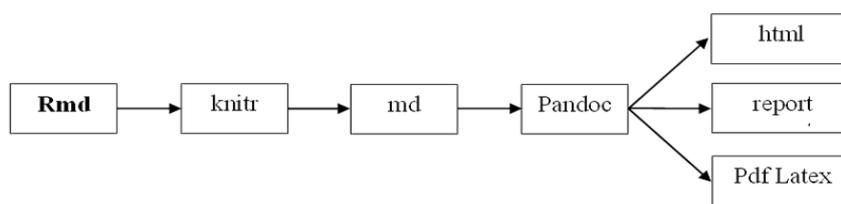
Markdown là ngôn ngữ đánh dấu văn bản đã có mặt từ rất lâu, với cú pháp đơn giản, dễ hiểu nên đã được nhiều người biết đến và được sử dụng khá phổ biến. Đây là loại ngôn ngữ đánh dấu văn bản được tạo ra vào năm 2004 bởi John Gruber. Một trong những mục tiêu cốt lõi của Markdown là cho phép dễ đọc hơn khi so sánh với các dạng trình bày đánh dấu văn bản khác. Điều này làm cho việc viết Markdown và đọc mã Markdown trở nên rất đơn giản như một văn bản thuần túy. Markdown được coi như là một công cụ sử dụng rộng rãi nhằm tạo ra tài liệu tự động, tái lập và chia sẻ được, như các báo cáo. Nó có thể tạo ra kết quả ở dạng thống kê số hoặc ở dạng tương tác, các định dạng Word, PDF, HTML và các định dạng khác [11].

Rmarkdown là sự kết hợp giữa Markdown và ngôn ngữ lập trình thống kê R, trong quá trình trình bày báo cáo khoa học giữa các nội dung lý thuyết, người làm nghiên cứu có thể đan xen vào đó các phép tính toán và được mô tả bằng các dòng code thực thi tính toán thống kê để đưa ra kết quả ở dạng số hoặc dạng bảng biểu, biểu đồ trực quan. Điều này giúp cho người đọc được tường minh và nắm bắt được dễ dàng hơn rất nhiều khi họ phải lần tìm những thông tin liên quan đến cơ sở lý thuyết, cách triển khai và áp dụng vào thực tế tính toán. Tất cả những nội dung đó đều có thể diễn giải tốt thông qua nghiên cứu khả lập và sử dụng Rmarkdown.

### 3. Kết hợp R và Markdown

#### 3.1. Quá trình hình thành tài liệu số

Một quy trình kết hợp ngôn ngữ lập trình tính toán thống kê R và Markdown để tạo ra tài liệu số có thể được mô phỏng bởi quá trình như Hình 2 [13], đây là quy trình thực hiện tạo ra một tài liệu số có sử dụng phương pháp nghiên cứu khả lập, các dạng file chứa các thông tin bao gồm văn bản (text), code thực thi của chương trình và kết quả đều được sử dụng trong quy trình này:



Hình 2. Sử dụng R và Markdown

- **RMarkdown**: là một biến thể của Markdown chỉ dành riêng cho ngôn ngữ R - nó cho phép người dùng soạn một tài liệu sử dụng Markdown để tạo chữ cũng như nhúng mã lập trình R và hiển thị kết quả đầu ra. Các tệp RMarkdown có đuôi '.Rmd'.

- **knitr**: Là một thư viện dành riêng cho R, thư viện này sẽ đọc các đoạn mã lệnh (code), thực thi chúng, và kết hợp chúng vào lại tài liệu. Các kết quả như số liệu thống kê, hình vẽ, biểu đồ trực quan sẽ được hiển thị kết hợp.

- **Markdown** không chỉ có thể kết hợp với riêng R mà còn có thể kết hợp với rất nhiều các ngôn ngữ khác. Các tệp được viết trong Markdown có đuôi '.md'.

- **Pandoc**: Cho phép chuyển đổi kết quả đầu ra thành các định dạng file như Word, PDF/Latex, PowerPoint, ... Đây là một phần mềm tách biệt ra khỏi ngôn ngữ R; tuy nhiên, nó có thể kết hợp với R một cách tự động thông qua môi trường tích hợp giao diện người dùng của R là Rstudio.

- Trong trường hợp chúng ta muốn xuất file văn bản hoặc báo cáo định dạng LaTeX, trước tiên cần cài đặt trình phiên dịch MikTeX, tiếp đó cài đặt gói Tinytex dành cho Rstudio. TinyTeX là một phiên bản LaTeX nhẹ, di động, đa nền tảng và dễ bảo trì, được hỗ trợ cho R trong việc xuất văn bản định dạng LaTeX. Để cài đặt TinyTeX, sử dụng lệnh: `install.packages('tinytex')`.

#### 3.2. Cấu trúc của một tài liệu số Rmarkdown

Về cơ bản, cấu trúc của một tài liệu Rmarkdown bao gồm 3 nội dung chính được miêu tả trong Hình 3.

- **YAML:** YAML là một ngôn ngữ tuần tự hóa dữ liệu thường được sử dụng để viết các tệp cấu hình. YAML rất phổ biến vì tính chất dễ đọc và dễ hiểu. YAML thường được kết hợp với các ngôn ngữ khác để tạo ra một quy trình tự động và tuần tự hóa theo các bước.

- **Code chunk:** Đây chính là nội dung mà chúng ta sẽ đưa các đoạn code tính toán thống kê hoặc xây dựng biểu đồ trực quan. Trong từng đoạn code chunk, dữ liệu sẽ được thực thi xử lý hoặc các phép toán thống kê mà nhà nghiên cứu sử dụng.

- **Markdown text:** Đây là các nội dung của Markdown, cho phép hiển thị dữ liệu dạng văn bản một cách linh hoạt, cho phép người đọc hiểu được cơ sở lý thuyết hoặc các nền tảng mà nhà nghiên cứu lựa chọn.

```

1 ---
2 title: "Understanding and hands-on with dumbbell charts"
3 author:
4   - Quang-Quy Tran[University of Information Technology and Communication, tqquy@ictu.edu.vn]
5   - Hong-Ngoc Dong[University of Economy and Business Administration, dhngoc@ueba.edu.vn]
6 date: "r format(sys.time(), '%d %B, %Y')%"
7 output: bookdown::gitbook
8 site: bookdown::bookdown_site
9 ---
10
11 ```{r setup, include=FALSE}
12 knitr::opts_chunk$set(echo = TRUE)
13 ```
14 |
15 # Introduction
16
17 In this paper, we are going to show you the definition and the goal of visualization by dumbbell
chart. Dumbbell charts are useful for displaying change between two time points for
several groups or observations. Using the gapminder dataset let's plot the change in life expectancy from
1952 to 2007 in the Asia. The dataset is in long format. We will need to convert it to wide format in order

```

Hình 3. Cấu trúc cơ bản của tài liệu Rmarkdown

### 3.3. Xây dựng tài liệu số

#### Tài liệu số dạng PDF LaTeX

Trong nội dung này, tiến hành tuần tự các bước sau để tạo báo cáo khoa học dạng LaTeX:

- Tạo ra một file Rmarkdown mới và thiết lập lần lượt các nội dung trong thẻ YAML bao gồm: Tên tiêu đề bài báo cáo, tên các tác giả nghiên cứu, thời gian xuất bản báo cáo, nội dung tóm tắt của báo cáo và các từ khóa định dạng đầu ra cho báo cáo.

- Trong Rstudio đã cài Tinytex và MikTex để hiển thị văn bản dạng Latex.

- Thêm nội dung: bibliography: *references.bib* và *csl: ieee-computational-intelligence-magazine* để trích dẫn tài liệu tham khảo dạng chuẩn IEEE tạp chí Computational intelligence, ở phần này nhóm tác giả sử dụng chuẩn IEEE tạp chí Computational intelligence để tạo trích dẫn tài liệu tham khảo làm ví dụ minh họa, ngoài ra các định dạng với các tạp chí khác hoàn toàn có thể sử dụng thông qua việc tải định dạng csl tại địa chỉ: <https://www.zotero.org/styles/>.

- *nocite: '@\*'*: Sử dụng để ghép tất cả các code chunks dành cho người đọc muốn theo dõi từ đầu báo cáo đến cuối báo cáo.

Nội dung này được miêu tả chi tiết khi cấu hình file .Rmd như Hình 4.

```

1 ---
2 title: |
3   |[[[logo.png]](width=0.8in)]
4 subtitle: "HỌC MÁY VỚI NGÔN NGỮ R"
5
6 author:
7   - TS. Trần Quang Quý[^BỘ môn Khoa học Máy tính, tqquy@ictu.edu.vn]
8 date: "Thái Nguyên, tháng 02 năm 2024"
9 geometry: "left=2.5cm,right=1.25cm,top=1.25cm,bottom=1.25cm"
10 output:
11   pdf_document:
12     fig_caption: true
13     number_sections: true
14     toc: true
15     toc_depth: 3
16     latex_engine: xelatex
17 header-includes:
18   - \usepackage[fontsize=11pt]{scrltendx}
19   - \usepackage[utf8]{inputenc}
20   - \usepackage{titlepic}
21   - \usepackage{caption}
22   - \captionsetup[figure]{name=Hình}
23   - \captionsetup[table]{name=Bảng}
24   - \renewcommand{\contentsname}{Mục lục}
25   # - \listoffigures
26   # - \listoftables
27   - \usepackage[nottoc]{tocbibind}
28
29 bibliography: references.bib
30 csl: ieee-computational-intelligence-magazine
31 nocite: '@*'

```

Hình 4. Cấu hình các gói và tham số cần thiết cho tài liệu số dạng PDF LaTeX

Kết quả thu được với tiêu đề, mục lục và nội dung văn bản kèm code minh họa như Hình 5 và Hình 6.



HỌC MÁY VỚI NGÔN NGỮ R  
TS. Trần Quang Quý\*  
Thái Nguyên, tháng 02 năm 2024

**Mục lục**

<b>1 Thao tác và biểu dữ liệu</b>	<b>2</b>
1.1 Cấu trúc dữ liệu trong R	2
1.1.1 Véc tơ	2
1.1.2 Factors	3
1.1.3 Danh sách	4
1.1.4 Data frames	5
1.1.5 Nhập dữ liệu và lưu trữ dạng CSV	7
1.2 Khám phá và biểu dữ liệu	7
1.2.1 Đơn vị đo xu hướng tập trung	8
1.2.2 Đơn vị đo hướng phân tán dữ liệu	9
1.2.3 Trục quan dữ liệu số	10
1.2.4 Biểu đồ histogram	11
1.2.5 Phân bố sai và độ lệch chuẩn	13
1.2.6 Dữ liệu định tính	15
1.2.7 Phân tích tương quan với scatter	15
1.2.8 Two-way cross tabulation	17
<b>2 Bộ phân loại KNN</b>	<b>18</b>
2.1 Độ hướng độ tương đồng	22
2.2 Lựa chọn số k	23
2.3 Chuẩn hóa dữ liệu	24
2.4 Case Study	25
2.4.1 Đọc dữ liệu	26
2.4.2 Chuẩn hóa dữ liệu	28
2.4.3 Phân tích dữ liệu	28
2.4.4 Hứa hẹn mô hình	29
2.4.5 Đánh giá mô hình	30
2.4.6 Cài thiện hiệu suất mô hình	30
<b>3 Tài liệu tham khảo</b>	<b>32</b>
4 Phụ lục: Phần tổng hợp các code Python và R được sử dụng	33

\*Đội môn Khoa học Máy tính, tqquy@ictu.edu.vn

Hình 5. File dạng PDF LaTeX được hiển thị

1.1.1 Véc tơ

Dạng cấu trúc dữ liệu cơ bản nhất của R được gọi là Véc tơ, nó lưu trữ một tập các giá trị thành phần (element). Một Véc tơ có thể lưu trữ các dữ liệu số, ký tự... Tuy nhiên cần nhớ rằng các thành phần phải chung 1 kiểu (same type) giá trị. Một véc tơ không thể chứa cả dữ liệu số và ký tự, thông thường để xác định dạng dữ liệu của véc tơ v, chúng ta sử dụng hàm typeof(v).

Một vài dạng véc tơ thường được dùng lưu trữ dữ liệu trong học máy bao gồm:

- integer - kiểu số nguyên
- double - kiểu số thực
- character - kiểu chuỗi
- logical - Giá trị TRUE hoặc FALSE
- NA - Giá trị bị thiếu (missing value)

Để tạo ra một véc tơ, chúng ta sử dụng ký hiệu c() (viết tắt của concatenate để nối dữ liệu vào với nhau). Một véc tơ có thể được đặt tên bằng cách sử dụng dấu mũi tên để gán (=). Toán tử này trong ngôn ngữ R được gọi là assignment operator.

Chúng ta cùng xem xét một vài ví dụ dưới đây tạo ra các véc tơ:

```
subject_name <- c("Quang Quý", "John", "Grigori")
temperature <- c(98.1, 98.6, 101.4)
flu_status <- c(FALSE, FALSE, TRUE)

# In ra các thành phần của subject_name
print(subject_name)
```

Để in và truy cập vào các thành phần của véc tơ, sử dụng chỉ số và dấu [].

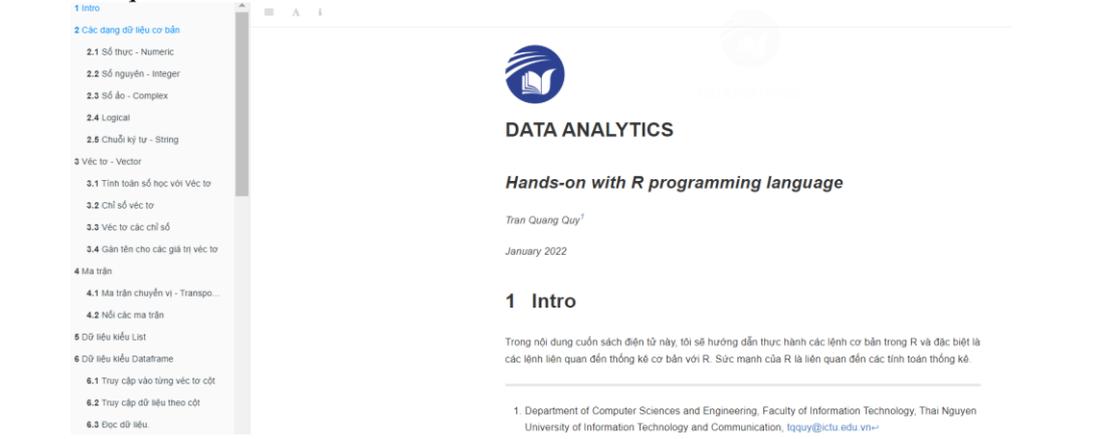
Hình 6. Dữ liệu text, code và kết quả cùng được hiển thị trong tài liệu

Tài liệu số dạng e-book

Để xuất bản được báo cáo dạng e-book, thực hiện các bước sau:

- Cài đặt thêm gói bookdown bằng lệnh: `install.packages("bookdown")`.
- Các thông số khác cho các thẻ cài đặt tương tự như với tài liệu PDFLatex.
- Thay đổi đầu ra của báo cáo với nội dung sau: `output: bookdown::gitbook; site: bookdown::bookdown_site`.

Khi đó, e-book được xuất bản ở dạng html và có thể được sử dụng như một quyển sách điện tử với các nội dung liên quan. Dạng sách e-book này có thể lưu trữ trên nhiều nền tảng và có thể sử dụng bất kỳ khi nào chỉ với một đường link, rất thuận tiện trong việc tra cứu, học và thực hành lập trình, kết quả thu được như Hình 7:



Hình 7. Hiển thị nội dung tài liệu dạng e-book

4. Kết luận

Mục đích của bài báo nhằm giới thiệu về ngôn ngữ R và Markdown trong việc tạo và xây dựng nguồn tài liệu số dành cho sinh viên chuyên ngành công nghệ thông tin. Phương pháp nghiên cứu khả lặp cho phép sinh viên tham gia tích cực trong quá trình học tập. Sinh viên có thể

thực hành và áp dụng kiến thức ngay trong quá trình học, thay vì chỉ lý thuyết trên bài vở. Điều này giúp củng cố hiểu biết và phát triển kỹ năng thực tế, việc rèn luyện kỹ năng lập trình trở nên cực kỳ quan trọng, đặc biệt khi thiếu đi sự hướng dẫn trực tiếp từ giáo viên.

Sử dụng tài liệu số cho phép sinh viên tiếp cận vào bất kỳ thời điểm nào và từ bất kỳ địa điểm nào có kết nối Internet. Không cần có sự hiện diện trực tiếp của giảng viên, sinh viên có thể tự học theo lịch trình riêng của mình. Điều này tiết kiệm thời gian và tạo điều kiện thuận lợi cho sinh viên làm việc theo sở thích và tiến độ cá nhân. Sinh viên có thể nghiên cứu từ các nguồn tài liệu chính thống, bài viết khoa học, tài liệu hướng dẫn từ các nhà cung cấp công nghệ và thậm chí cả các dự án mã nguồn mở. Điều này giúp mở rộng kiến thức và đa dạng hóa phạm vi học tập.

Tóm lại, phương pháp nghiên cứu khả lập trong việc xây dựng tài liệu số trong giảng dạy Công nghệ thông tin tạo ra môi trường học tập tương tác, linh hoạt và thú vị cho sinh viên. Nó khuyến khích sự tự chủ và phát triển kỹ năng thực tế, đồng thời mang lại sự tiết kiệm thời gian và tiếp cận tài liệu đa dạng.

#### TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] T. Q. P. Thieu and M. V. Nguyen, "Solutions to improve programming skills for information technology students, An Giang University," *Vietnam Journal of Education*, no. 473, pp. 40-44, 2020.
- [2] T. T. Phan, "Developing information technology competence for students majored in Primary education at Pedagogical university to meet the requirements of educational innovation," *Vietnam Journal of Educational Sciences*, no. 42, pp. 28-33, 2021.
- [3] T. T. T. Nguyen, "Applying information technology in higher education," *Trade union Scientific Research magazine*, no. 27, pp. 46-51, September 2022.
- [4] L. U. M. Huynh and H. T. Nguyen, "Group learning solution for students majoring in Information Technology, Dong Thap University," *Journal of Scientific Research of Dong Thap University*, no. 9, pp. 67-81, 2020.
- [5] T. B. Nguyen, "Improving capacity of information technology application in experience teaching for teachers and students at Hanoi metropolitan university," *Vietnam Journal of Education*, no. 69, pp. 63-73 February 2023.
- [6] T. M. D. Nguyen and T. T. T. Nguyen, "Solutions to improve professional capacity for information technology lecturers in the context of the 4.0 industrial revolution at Dong Thap University," *Journal of Scientific Research of Dong Thap University*, no. 9, pp. 57-66, 2020.
- [7] M. J. T. Milton and A. Possolo, "Trustworthy data underpin reproducible research," *Nature Physics*, vol. 16, no. 2, pp. 114-119, 2020.
- [8] K. Peter, J. Holler, and S. Bardin, "Reproducible Research Practices and Barriers to Reproducible Research in Geography: Insights from a Survey," *Annals of the American Association of Geographers*, vol. 114, no. 2, pp. 369-386, 2023.
- [9] R. D. Peng, "Reproducible research and biostatistics," *Biostatistics*, vol. 10, no. 3, pp. 405-408, 2009.
- [10] P. Aaron, C. J. Van Lissa, and A. M. Brandmaier, "Reproducible research in R: A tutorial on how to do the same thing more than once," *MDPI Journal*, vol. 3, no. 4, pp. 836-867, 2021.
- [11] B. Benjamin and D. Udwin, "R markdown," *Computational Statistics*, vol. 7, no. 3, pp. 167-177, 2015.