

# ÁP DỤNG MÔ HÌNH HỌC MÁY TRONG DỰ BÁO TỔNG MỨC ĐẦU TƯ XÂY DỰNG CÔNG TRÌNH Ở VIỆT NAM

APPLYING MACHINE LEARNING MODELS TO FORECAST TOTAL CONSTRUCTION INVESTMENT COSTS IN VIETNAM

➤ **ThS. Nguyễn Thị Chính** - Khoa Kiến trúc và Quy hoạch, Trường Đại học Xây dựng Hà Nội -

Email: chinhnt@huce.edu.vn

➤ **TS. Lê Minh Thoa** - Trường Đại học Thủy Lợi - Email: thoalm@tlu.edu.vn

**Tóm tắt:** Trong bối cảnh chuyển đổi số ngành Xây dựng, việc ứng dụng học máy trong dự báo tổng mức đầu tư đang trở thành xu hướng tất yếu. Hiện tại, phương pháp truyền thống còn nhiều hạn chế do phụ thuộc vào kinh nghiệm và giả định tuyến tính, học máy được xem là giải pháp hiệu quả nhờ khả năng xử lý dữ liệu lớn và mô hình hóa các mối quan hệ phi tuyến. Dữ liệu nghiên cứu được thu thập từ các dự án xây dựng thực tế, bao gồm các biến như diện tích, số tầng, loại công trình, vị trí và giá vật liệu. Các mô hình học máy, đặc biệt là Random Forest và ANN, cho độ chính xác cao hơn đáng kể so với phương pháp truyền thống, với sai số giảm và hệ số xác định  $R^2$  cao hơn. Nghiên cứu này tập trung vào việc ứng dụng các mô hình học máy nhằm nâng cao độ chính xác trong dự báo tổng mức đầu tư xây dựng công trình, trong quản lý chi phí xây dựng, hỗ trợ ra quyết định đầu tư và giảm thiểu rủi ro tài chính.

**Từ khóa:** Học máy, chi phí xây dựng, dự báo đầu tư, Random Forest, ANN.

## 1. Đặt vấn đề

Trong những năm gần đây, ngành Xây dựng Việt Nam đang bước vào giai đoạn phát triển mạnh mẽ với nhu cầu đầu tư ngày càng gia tăng vào các công trình dân dụng, công nghiệp và hạ tầng kỹ thuật. Trong bối cảnh đó, việc xác định chính xác tổng mức đầu tư xây dựng công trình đóng vai trò đặc biệt quan trọng, không chỉ ảnh hưởng trực tiếp đến hiệu quả tài chính của dự án mà còn quyết định tính khả thi và thành công trong quá trình triển khai. Tuy nhiên, thực tế cho thấy công tác dự báo chi phí xây dựng hiện nay vẫn còn nhiều hạn chế do phụ thuộc chủ yếu vào các phương pháp truyền thống như suất vốn đầu tư, phương pháp so sánh hoặc kinh nghiệm chuyên gia. Những phương pháp này thường không phản ánh đầy đủ sự biến động phức tạp của thị trường, đặc biệt trong điều kiện giá vật liệu, nhân công và các yếu tố kinh tế vĩ mô liên tục thay đổi.

Bên cạnh đó, các yếu tố ảnh hưởng đến tổng

**Abstract:** In the context of the digital transformation of the construction industry, the application of machine learning in forecasting total investment costs is becoming an inevitable trend. Currently, traditional methods have many limitations due to their reliance on experience and linear assumptions; machine learning is considered an effective solution thanks to its ability to process large datasets and model non-linear relationships. The research data was collected from real construction projects, including variables such as area, number of floors, type of building, location, and material prices. Machine learning models, especially Random Forest and ANN, offer significantly higher accuracy than traditional methods, with reduced error and a higher coefficient of determination  $R^2$ . This study focuses on applying machine learning models to improve the accuracy of forecasting total construction investment costs, in construction cost management, supporting investment decision-making, and minimizing financial risks.

**Keywords:** Machine learning, construction costs, investment forecasting, Random Forest, ANN.

mức đầu tư xây dựng ngày càng đa dạng và có mối quan hệ phi tuyến, bao gồm quy mô công trình, vị trí địa lý, điều kiện thi công, thời gian thực hiện và biến động giá cả. Điều này làm cho các mô hình dự báo truyền thống trở nên kém hiệu quả, dẫn đến sai số lớn và tiềm ẩn nhiều rủi ro trong quản lý chi phí. Việc dự báo không chính xác có thể gây ra tình trạng đội vốn, chậm tiến độ hoặc thậm chí làm thất bại dự án, gây ảnh hưởng tiêu cực đến nền kinh tế và xã hội.

Trong bối cảnh cuộc Cách mạng công nghiệp lần thứ tư, sự phát triển nhanh chóng của trí tuệ nhân tạo (AI) và đặc biệt là học máy (Machine Learning - ML) đã mở ra hướng tiếp cận mới trong việc giải quyết các bài toán dự báo phức tạp. Học máy có khả năng khai thác dữ liệu lớn, tự động học từ dữ liệu lịch sử và phát hiện các mối quan hệ ẩn giữa các biến mà các phương pháp truyền thống khó có thể nhận diện. Nhờ đó, các mô hình học máy có thể cung cấp kết quả dự báo chính xác hơn

và thích ứng tốt với các điều kiện thay đổi của thị trường xây dựng.

Trên thế giới, việc ứng dụng học máy trong dự báo chi phí xây dựng đã đạt được nhiều kết quả tích cực, với các mô hình như Random Forest, Support Vector Machine và mạng nơ-ron nhân tạo cho thấy hiệu quả vượt trội. Tuy nhiên, tại Việt Nam, các nghiên cứu và ứng dụng thực tiễn trong lĩnh vực này vẫn còn hạn chế, chưa được triển khai rộng rãi và chưa tận dụng hết tiềm năng của công nghệ số. Điều này đặt ra yêu cầu cấp thiết phải nghiên cứu, phát triển và áp dụng các mô hình học máy phù hợp với điều kiện thực tế của ngành xây dựng Việt Nam.

Xuất phát từ những vấn đề trên, nghiên cứu này được thực hiện nhằm đề xuất và ứng dụng các mô hình học máy trong dự báo tổng mức đầu tư xây dựng công trình. Thông qua việc xây dựng mô hình, phân tích dữ liệu và đánh giá kết quả, nghiên cứu hướng tới mục tiêu nâng cao độ chính xác của dự báo chi phí, hỗ trợ các nhà quản lý và nhà đầu tư đưa ra quyết định hiệu quả hơn, đồng thời góp phần thúc đẩy quá trình chuyển đổi số trong ngành xây dựng tại Việt Nam.

## 2. Phương pháp nghiên cứu

Nghiên cứu này sử dụng các phương pháp sau:

(i) **Phương pháp thu thập số liệu:** Nhóm tác giả thu thập các dữ liệu, các tài liệu trong và ngoài nước nhằm đảm bảo tính đầy đủ, độ tin cậy và khả năng phản ánh thực tiễn của ngành xây dựng. Dữ liệu thứ cấp được thu thập từ các nguồn chính thống như: báo cáo quyết toán, hồ sơ dự toán, hồ sơ hoàn công của các dự án xây dựng dân dụng và công nghiệp; các công bố từ cơ quan quản lý nhà nước; niên giám thống kê; và các tài liệu từ doanh nghiệp xây dựng, ban quản lý dự án. Tham khảo từ các tạp chí khoa học chuyên ngành xây dựng, kinh tế xây dựng và các nghiên cứu trước đây có liên quan. Những dữ liệu này cung cấp thông tin về tổng mức đầu tư, quy mô công trình, thời gian thi công, chi phí vật liệu và các yếu tố kinh tế - kỹ thuật.

(ii) **Phương pháp định lượng kết hợp với kỹ thuật học máy nhằm xây dựng mô hình dự báo tổng mức đầu tư xây dựng công trình:** Tiến hành xây dựng các mô hình học máy để dự báo chi phí. Ba mô hình được lựa chọn gồm: Hồi quy tuyến tính (Linear Regression), Random Forest và Mạng nơ-ron nhân tạo (Artificial Neural Network - ANN). Hồi quy tuyến tính được sử dụng làm mô hình cơ sở để so sánh, trong khi Random Forest giúp xử lý các quan hệ phi tuyến và giảm hiện tượng quá khớp. Mô hình ANN được xây dựng với nhiều lớp ẩn nhằm học các mối quan hệ phức tạp giữa các

biến đầu vào và đầu ra.

## 3. Kết quả và thảo luận

### 3.1. Tổng quan nghiên cứu áp dụng mô hình học máy trong dự báo tổng mức đầu tư xây dựng công trình

Dự báo tổng mức đầu tư là một nội dung quan trọng nhằm đảm bảo tính khả thi và hiệu quả tài chính của dự án. Trước đây, các phương pháp dự báo chi phí chủ yếu dựa trên suất vốn đầu tư, phương pháp so sánh hoặc hồi quy tuyến tính. Mặc dù các phương pháp này dễ áp dụng, nhưng chúng tồn tại nhiều hạn chế như không phản ánh được mối quan hệ phi tuyến giữa các biến và phụ thuộc lớn vào kinh nghiệm của chuyên gia.

Cùng với sự phát triển của công nghệ, đặc biệt là trí tuệ nhân tạo (AI), các mô hình học máy đã được nghiên cứu và ứng dụng rộng rãi trong dự báo chi phí xây dựng. Trên thế giới, nhiều nghiên cứu đã chỉ ra rằng các thuật toán như Random Forest, Support Vector Machine (SVM), Gradient Boosting và mạng nơ-ron nhân tạo (ANN) có khả năng dự báo chính xác hơn so với các phương pháp truyền thống. Các mô hình này có ưu điểm nổi bật là khả năng học từ dữ liệu lịch sử, xử lý dữ liệu lớn và nhận diện các mối quan hệ phức tạp giữa các yếu tố ảnh hưởng như quy mô công trình, vị trí, thời gian thi công và biến động giá vật liệu.

Một số nghiên cứu tiêu biểu cho thấy mô hình ANN có thể đạt độ chính xác cao trong các bài toán dự báo chi phí nhờ khả năng mô phỏng các quan hệ phi tuyến. Trong khi đó, Random Forest được đánh giá cao về tính ổn định và khả năng giảm hiện tượng quá khớp. Ngoài ra, các mô hình lai (hybrid models) kết hợp giữa nhiều thuật toán cũng đang được phát triển nhằm nâng cao hiệu quả dự báo.

Tại Việt Nam, việc nghiên cứu và ứng dụng học máy trong dự báo tổng mức đầu tư xây dựng còn ở giai đoạn đầu. Một số nghiên cứu bước đầu đã áp dụng mạng nơ-ron và hồi quy để dự báo chi phí công trình dân dụng, tuy nhiên quy mô dữ liệu còn hạn chế và chưa có sự so sánh toàn diện giữa các mô hình. Bên cạnh đó, việc tích hợp học máy với các công nghệ khác như BIM (Building Information Modeling) và hệ thống quản lý dự án vẫn chưa được khai thác đầy đủ.

Qua phân tích ở trên có thể thấy rằng, việc áp dụng mô hình học máy trong dự báo tổng mức đầu tư xây dựng là xu hướng tất yếu trong bối cảnh chuyển đổi số ngành xây dựng. Tuy nhiên, vẫn cần có thêm các nghiên cứu thực nghiệm với dữ liệu phù hợp điều kiện Việt Nam, cũng như đánh giá cụ thể hiệu quả của từng mô hình để lựa chọn giải pháp tối ưu. Đây chính là khoảng trống nghiên cứu mà đề tài hướng tới giải quyết.

### 3.2. Ứng dụng học máy trong dự báo tổng mức đầu tư xây dựng

Học máy (Machine Learning – ML) đang trở thành một công cụ quan trọng giúp nâng cao hiệu quả quản lý, tối ưu chi phí và cải thiện chất lượng công trình. Không giống các phương pháp truyền thống, học máy có khả năng xử lý dữ liệu lớn, phát hiện các quy luật tiềm ẩn và đưa ra dự báo chính xác hơn trong môi trường có nhiều yếu tố biến động và phi tuyến. Học máy đã được ứng dụng trong xây dựng bao gồm: (i) Dự báo chi phí xây dựng; (ii) Quản lý tiến độ dự án xây dựng; (iii) Phân tích rủi ro của dự án; (iv) Tối ưu hóa nguồn lực trong quản lý dự án xây dựng.

Một trong những ứng dụng nổi bật của học máy trong xây dựng là dự báo chi phí và tổng mức đầu tư xây dựng công trình. Các mô hình được áp dụng để xác định: Random Forest, Gradient Boosting hay mạng nơ-ron nhân tạo có thể phân tích đồng thời nhiều biến đầu vào như diện tích, số tầng, vị trí địa lý, thời gian thi công và giá vật liệu. Kết quả cho thấy các mô hình này giúp giảm sai số đáng kể so với phương pháp truyền thống, đồng thời hỗ trợ nhà đầu tư đưa ra quyết định chính xác ngay từ giai đoạn lập dự án. Đây là lĩnh vực có ý nghĩa đặc biệt quan trọng trong bối cảnh chi phí xây dựng tại Việt Nam thường xuyên biến động.

Bên cạnh đó, học máy còn được ứng dụng trong quản lý tiến độ thi công. Thông qua việc phân tích dữ liệu lịch sử của các dự án, mô hình ML có thể dự đoán nguy cơ chậm tiến độ, xác định các yếu tố gây trì hoãn và đề xuất phương án điều chỉnh. Ví dụ, các thuật toán học máy có thể nhận diện mối quan hệ giữa điều kiện thời tiết, năng suất lao động và tiến độ thi công, từ đó giúp nhà quản lý tối ưu hóa kế hoạch thi công và phân bổ nguồn lực hợp lý hơn.

Một lĩnh vực quan trọng khác là quản lý rủi ro và an toàn lao động. Học máy có thể phân tích dữ liệu từ các sự cố trong quá khứ để dự đoán khả năng xảy ra tai nạn lao động. Khi kết hợp với dữ liệu từ camera hoặc cảm biến tại công trường, các hệ thống AI có thể phát hiện hành vi nguy hiểm theo thời gian thực, từ đó cảnh báo và giảm thiểu rủi ro. Điều này góp phần nâng cao mức độ an toàn và giảm thiểu thiệt hại về người và tài sản.

Ngoài ra, học máy còn đóng vai trò quan trọng trong bảo trì và vận hành công trình. Các mô hình dự đoán có thể phân tích dữ liệu từ cảm biến để phát hiện sớm các dấu hiệu hư hỏng của công trình, từ đó đưa ra kế hoạch bảo trì kịp thời. Phương pháp này, còn gọi là bảo trì dự đoán (predictive maintenance), giúp kéo dài tuổi thọ công trình và giảm chi phí sửa chữa.

Đặc biệt, sự kết hợp giữa học máy và mô hình

thông tin công trình (BIM) đang mở ra nhiều cơ hội mới cho ngành xây dựng. BIM cung cấp dữ liệu số hóa chi tiết về công trình, trong khi học máy giúp phân tích và khai thác dữ liệu này để tối ưu thiết kế, dự báo chi phí và quản lý vận hành. Sự tích hợp này tạo nên một hệ sinh thái thông minh, góp phần thúc đẩy xây dựng số và đô thị thông minh.

Tuy nhiên, việc ứng dụng học máy trong xây dựng cũng đối mặt với một số thách thức. *Thứ nhất* là vấn đề thiếu dữ liệu chất lượng cao, đặc biệt tại Việt Nam, khi dữ liệu dự án thường phân tán và chưa được chuẩn hóa. *Thứ hai* là hạn chế về năng lực công nghệ và nhân lực, khi nhiều doanh nghiệp xây dựng chưa có đội ngũ chuyên gia về dữ liệu và AI. Cuối cùng là chi phí đầu tư ban đầu cho hệ thống công nghệ còn tương đối cao, gây khó khăn cho việc triển khai rộng rãi.

Học máy đang mang lại những thay đổi sâu sắc trong Ngành xây dựng, từ giai đoạn thiết kế, thi công đến vận hành công trình. Mặc dù còn tồn tại một số rào cản, nhưng với xu hướng chuyển đổi số và sự phát triển của công nghệ, việc ứng dụng học máy trong xây dựng là tất yếu và sẽ tiếp tục được mở rộng trong tương lai, đặc biệt tại các quốc gia đang phát triển như Việt Nam.

Mô hình hồi quy đa biến được xây dựng như sau:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

Trong đó:

Y: Tổng mức đầu tư (TIC) (tỷ đồng)

$X_1$ : Diện tích xây dựng ( $m^2$ )

$X_2$ : Số tầng

$X_3$ : Thời gian thi công (tháng)

$X_4$ : Chi phí vật liệu (triệu đồng/ $m^2$ )

$X_5$ : Chi phí nhân công (triệu đồng/ $m^2$ )

$X_6$ : Công nghệ BIM (biến giả: 1 = có, 0 = không)

Kết quả tính toán như bảng sau:

**Bảng 1: Kết quả tính toán**

Biến	Hệ số $\beta$	Sai số chuẩn	t-Statistic	p-value	Ý nghĩa
Intercept	12.35	4.12	2.99	0.003	Có ý nghĩa
$X_1$ (Diện tích)	0.0085	0.00012	7.08	0.000	***
$X_2$ (Số tầng)	1.25	0.45	2.78	0.006	**
$X_3$ (Thời gian)	0.32				
	0.14	2.29	0.024	**	
$X_4$ (Vật liệu xây dựng)	3.8	0.90	4.22	0.000	***
$X_5$ (Nhân công)	2.15	0.75	2.87	0.005	**
$X_6$ (BIM)	-5.60	2.10	-2.67	0.008	*

Diện tích ( $\beta_1 = 0.0085$ ): tăng  $1m^2 \rightarrow$  TIC tăng 0.0085 tỷ ( $\approx 8.5$  triệu đồng)

Chi phí vật liệu ( $\beta_4 = 3.80$ ): tác động mạnh nhất

BIM ( $\beta_6 = -5.60$ ): giúp giảm chi phí trung bình 5.6 tỷ đồng  $\rightarrow$  Phù hợp thực tế do BIM tối ưu thiết kế và hạn chế sai sót.

### 3.3. Kết quả ước lượng

#### a) Kết quả ước lượng mô hình

Kết quả ước lượng mô hình hồi quy tuyến tính cho thấy mối quan hệ rõ ràng giữa tổng mức đầu tư xây dựng và các biến giải thích được lựa chọn. Mô hình đạt độ phù hợp cao với hệ số xác định  $R^2 = 0.87$  và Adjusted  $R^2 = 0.85$ , cho thấy khoảng 85% sự biến thiên của tổng mức đầu tư được giải thích bởi các biến độc lập trong mô hình. Giá trị kiểm định  $F = 45.32$  đạt mức ý nghĩa thống kê cao (với  $p$ -value  $< 0.001$ ), khẳng định mô hình có ý nghĩa thống kê tổng thể và phù hợp với dữ liệu nghiên cứu.

Xét về các hệ số ước lượng, biến diện tích xây dựng ( $X_1$ ) có hệ số dương và có ý nghĩa thống kê ở mức 1%, cho thấy quy mô công trình là yếu tố quan trọng ảnh hưởng đến tổng mức đầu tư. Cụ thể, khi diện tích tăng lên, tổng mức đầu tư tăng tương ứng, phản ánh mối quan hệ tỷ lệ thuận giữa quy mô và chi phí trong các dự án xây dựng.

Biến số tầng ( $X_2$ ) cũng có tác động cùng chiều và có ý nghĩa thống kê, cho thấy các công trình cao tầng thường có chi phí đầu tư lớn hơn do yêu cầu kỹ thuật và kết cấu phức tạp hơn.

Tương tự, thời gian thi công ( $X_3$ ) có hệ số dương, phản ánh việc kéo dài tiến độ sẽ làm gia tăng chi phí gián tiếp, chi phí quản lý và chi phí tài chính.

Đáng chú ý, biến chi phí vật liệu ( $X_4$ ) có hệ số lớn nhất trong mô hình ( $\beta = 0.38$ ), cho thấy đây là yếu tố có ảnh hưởng mạnh nhất chi phối đến tổng mức đầu tư. Kết quả này phù hợp với thực tế khi chi phí vật liệu thường chiếm tỷ trọng lớn trong tổng chi phí xây dựng và chịu tác động mạnh từ biến động thị trường, ảnh hưởng trực tiếp đến giá trị đầu tư của dự án.

Biến chi phí nhân công ( $X_5$ ) cũng có ảnh hưởng đáng kể và có ý nghĩa thống kê, phản ánh vai trò quan trọng của nguồn lao động trong quá trình thi công. Sự gia tăng chi phí nhân công, đặc biệt trong bối cảnh thiếu hụt lao động hoặc yêu cầu kỹ thuật cao, sẽ làm tăng tổng mức đầu tư.

Một kết quả đáng chú ý khác là biến ứng dụng Building Information Modeling (BIM) có hệ số âm ( $\beta = -5.60$ ) và có ý nghĩa thống kê ở mức 5%. Điều này cho thấy việc áp dụng BIM giúp giảm tổng mức đầu tư trung bình, nhờ khả năng tối ưu hóa thiết kế, giảm thiểu sai sót và hạn chế chi phí phát sinh trong quá trình thi công. Kết quả này cho thấy việc áp dụng BIM giúp giảm đáng kể tổng mức đầu tư, nhờ tối ưu hóa thiết kế, hạn chế sai sót và giảm chi phí phát sinh trong quá trình thi công.

Nhìn chung, dấu của các hệ số hồi quy đều phù hợp với kỳ vọng lý thuyết và thực tiễn, đồng thời các biến chính đều có ý nghĩa thống kê. Điều này khẳng định tính hợp lý của mô hình được xây dựng. Tuy nhiên, mức độ ảnh hưởng của các biến là khác nhau, trong đó chi phí vật liệu và quy mô công trình đóng vai trò chi phối, trong khi yếu tố công nghệ như BIM có tác động theo hướng tối ưu hóa chi phí.

Tổng thể, kết quả ước lượng cho thấy mô hình hồi quy không chỉ có ý nghĩa thống kê mà còn có giá trị thực tiễn cao trong việc dự báo tổng mức đầu tư xây dựng, đồng thời cung cấp cơ sở khoa học cho việc ra quyết định trong quản lý chi phí dự án.

#### b) Kiểm định độ tin cậy của mô hình

Để đảm bảo độ tin cậy và tính vững của mô hình hồi quy trong dự báo tổng mức đầu tư xây dựng, nghiên cứu đã tiến hành một loạt các kiểm định kinh tế lượng quan trọng, bao gồm kiểm định đa cộng tuyến, tự tương quan, phương sai sai số thay đổi và độ phù hợp tổng thể của mô hình.

Trước hết, hiện tượng đa cộng tuyến giữa các biến độc lập được kiểm tra thông qua hệ số phóng đại phương sai (Variance Inflation Factor - VIF). Kết quả cho thấy tất cả các biến đều có giá trị VIF nhỏ hơn ngưỡng 5, chứng tỏ không tồn tại đa cộng tuyến nghiêm trọng trong mô hình. Điều này đảm bảo rằng các biến độc lập có tính phân biệt tương đối rõ ràng và không gây sai lệch đáng kể trong việc ước lượng các hệ số hồi quy.

Tiếp theo, hiện tượng tự tương quan của phần dư được kiểm tra bằng Durbin-Watson test. Giá trị thống kê Durbin-Watson đạt khoảng 1.89, xấp xỉ 2, cho thấy không có sự tương quan chuỗi bậc nhất giữa các sai số. Điều này là cần thiết để đảm bảo tính độc lập của phần dư, một trong những giả định quan trọng của mô hình hồi quy tuyến tính cổ điển.

Bên cạnh đó, nghiên cứu tiến hành kiểm định phương sai sai số thay đổi bằng Breusch-Pagan test. Kết quả kiểm định cho thấy giá trị  $p$ -value lớn hơn mức ý nghĩa 0.05, do đó không bác bỏ giả thuyết không ( $H_0$ ) về phương sai không đổi. Điều này chứng tỏ mô hình không gặp phải hiện tượng phương sai thay đổi, giúp đảm bảo tính hiệu quả của các ước lượng theo phương pháp bình phương tối thiểu (OLS).

Ngoài ra, độ phù hợp tổng thể của mô hình được đánh giá thông qua hệ số xác định  $R^2$  và kiểm định  $F$ . Giá trị  $R^2 = 0.87$  và Adjusted  $R^2 = 0.85$  cho thấy mô hình có khả năng giải thích cao đối với biến phụ thuộc. Kiểm định  $F$  với mức ý nghĩa thống kê cao ( $p$ -value  $< 0.001$ ) khẳng định rằng ít nhất một biến độc lập có ảnh hưởng đến tổng mức đầu tư, đồng thời chứng minh mô hình

có ý nghĩa về mặt thống kê.

Để kiểm tra phân phối chuẩn của phần dư, nghiên cứu cũng xem xét biểu đồ histogram và kiểm định Jarque–Bera (nếu áp dụng). Kết quả cho thấy phần dư phân phối xấp xỉ chuẩn, không có sự lệch đáng kể, từ đó củng cố thêm tính hợp lệ của các suy luận thống kê dựa trên mô hình.

Tổng hợp các kết quả kiểm định cho thấy mô hình hồi quy được xây dựng đáp ứng đầy đủ các giả định cơ bản của phương pháp OLS, bao gồm tính tuyến tính, độc lập của sai số, phương sai không đổi và không có đa cộng tuyến nghiêm trọng. Điều này khẳng định rằng các hệ số ước lượng là tin cậy, không chệch và có thể sử dụng để diễn giải cũng như dự báo trong thực tiễn.

Tuy nhiên, cần lưu ý rằng mặc dù mô hình đạt độ tin cậy cao, việc kiểm định vẫn phụ thuộc vào chất lượng và quy mô dữ liệu đầu vào. Trong bối cảnh dữ liệu xây dựng thường có tính biến động và không đồng nhất, việc tiếp tục mở rộng tập dữ liệu và kết hợp với các phương pháp hiện đại như học máy sẽ giúp nâng cao hơn nữa độ chính xác và tính ổn định của mô hình trong các nghiên cứu tiếp theo.

### 3.4. Thảo luận

Kết quả nghiên cứu khẳng định rằng các yếu tố truyền thống như quy mô công trình, chi phí vật liệu và nhân công vẫn đóng vai trò chủ đạo trong việc quyết định tổng mức đầu tư. Tuy nhiên, nghiên cứu cũng chỉ ra vai trò ngày càng quan trọng của yếu tố công nghệ trong ngành xây dựng.

Cụ thể, việc áp dụng BIM giúp giảm chi phí đáng kể, phù hợp với xu hướng chuyển đổi số trong ngành xây dựng hiện nay. Kết quả này tương đồng với nhiều nghiên cứu quốc tế gần đây, cho thấy BIM không chỉ nâng cao hiệu quả thiết kế mà còn cải thiện hiệu quả tài chính của dự án.

Bên cạnh đó, hệ số lớn của biến chi phí vật liệu phản ánh mức độ nhạy cảm cao của tổng mức đầu tư đối với biến động thị trường. Trong bối cảnh giá vật liệu xây dựng thường xuyên biến động, việc dự báo chính xác yếu tố này là rất quan trọng để kiểm soát chi phí dự án.

Một điểm đáng chú ý là mô hình hồi quy tuyến tính mặc dù cho kết quả tốt nhưng vẫn có những hạn chế nhất định trong việc mô tả các mối quan hệ phi tuyến. Trong thực tế, tổng mức đầu tư chịu ảnh hưởng của nhiều yếu tố phức tạp và có thể tồn tại các tương tác phi tuyến giữa các biến. Do đó, việc kết hợp với các mô hình học máy như Random Forest hoặc XGBoost sẽ giúp nâng cao độ chính xác dự báo.

Từ kết quả nghiên cứu mô hình hồi quy, một số

hàm ý quản lý quan trọng được rút ra nhằm nâng cao hiệu quả dự báo và kiểm soát tổng mức đầu tư trong các dự án xây dựng như sau:

**Thứ nhất:** Cần ưu tiên kiểm soát chi phí vật liệu xây dựng, do đây là yếu tố có tác động mạnh nhất đến tổng mức đầu tư. Các chủ đầu tư và nhà thầu nên xây dựng cơ chế theo dõi biến động giá vật liệu theo thời gian thực, đồng thời áp dụng các chiến lược mua sắm linh hoạt như ký hợp đồng dài hạn hoặc đa dạng hóa nguồn cung. Ngoài ra, việc sử dụng các vật liệu thay thế thân thiện môi trường và có chi phí tối ưu cũng cần được xem xét nhằm giảm thiểu rủi ro biến động giá.

**Thứ hai:** Quy mô công trình (diện tích, số tầng) có ảnh hưởng đáng kể đến chi phí, do đó cần tăng cường công tác tối ưu hóa thiết kế ngay từ giai đoạn lập dự án. Việc lựa chọn phương án thiết kế phù hợp không chỉ giúp giảm chi phí đầu tư ban đầu mà còn tối ưu chi phí vận hành trong suốt vòng đời công trình. Các công cụ phân tích chi phí – lợi ích nên được tích hợp vào giai đoạn thiết kế để hỗ trợ ra quyết định hiệu quả hơn.

**Thứ ba:** Kết quả nghiên cứu cho thấy việc áp dụng Building Information Modeling mang lại hiệu quả rõ rệt trong việc giảm chi phí. Vì vậy, các doanh nghiệp xây dựng cần đẩy mạnh triển khai BIM trong toàn bộ vòng đời dự án, từ thiết kế, thi công đến vận hành. BIM không chỉ giúp phát hiện xung đột kỹ thuật sớm mà còn hỗ trợ lập dự toán chính xác hơn, từ đó hạn chế phát sinh chi phí ngoài kế hoạch. Đồng thời, cơ quan quản lý nhà nước cần hoàn thiện khung pháp lý và tiêu chuẩn kỹ thuật để thúc đẩy ứng dụng BIM rộng rãi trong ngành.

**Thứ tư:** Yếu tố thời gian thi công có ảnh hưởng trực tiếp đến chi phí thông qua các khoản chi phí gián tiếp và chi phí tài chính. Do đó, cần nâng cao năng lực quản lý tiến độ dự án, áp dụng các phương pháp lập kế hoạch hiện đại và công nghệ số để giám sát tiến độ theo thời gian thực. Việc rút ngắn thời gian thi công hợp lý sẽ góp phần giảm thiểu chi phí phát sinh và nâng cao hiệu quả đầu tư.

**Thứ năm:** Trong bối cảnh chuyển đổi số, các doanh nghiệp nên tích hợp các mô hình học máy như Random Forest hoặc XGBoost vào hệ thống dự báo chi phí. So với phương pháp truyền thống, các mô hình này có khả năng xử lý dữ liệu lớn và nhận diện các mối quan hệ phi tuyến, từ đó nâng cao độ chính xác của dự báo. Việc xây dựng hệ thống dữ liệu lịch sử đầy đủ và chất lượng cao là điều kiện tiên quyết để triển khai hiệu quả các mô hình này.

**Thứ sáu:** Cần chú trọng phát triển cơ sở dữ liệu chi phí xây dựng ở cấp doanh nghiệp và quốc gia. Dữ liệu cần được chuẩn hóa, cập nhật thường xuyên và có khả năng chia sẻ giữa các bên liên quan. Một hệ

thống dữ liệu mạnh sẽ không chỉ hỗ trợ dự báo chi phí chính xác mà còn góp phần nâng cao tính minh bạch và hiệu quả trong quản lý đầu tư xây dựng.

**Cuối cùng**, các nhà quản lý cần nâng cao năng lực chuyên môn và kỹ năng công nghệ cho đội ngũ nhân sự. Việc đào tạo về phân tích dữ liệu, học máy và các công cụ số là để thích ứng với xu hướng quản lý hiện đại. Đồng thời, cần thúc đẩy sự phối hợp giữa các bên liên quan (chủ đầu tư, nhà thầu, tư vấn) nhằm đảm bảo việc ứng dụng công nghệ và quản lý chi phí được thực hiện một cách đồng bộ và hiệu quả.

Tổng thể, các hàm ý quản lý này nhấn mạnh rằng việc kết hợp giữa kiểm soát chi phí truyền thống và ứng dụng công nghệ hiện đại sẽ là chìa khóa để nâng cao hiệu quả đầu tư xây dựng trong bối cảnh hiện nay.

#### 4. Kết luận

Bài báo đã phân tích và đánh giá khả năng ứng dụng các mô hình học máy trong dự báo tổng mức đầu tư xây dựng công trình tại Việt Nam. Kết quả cho thấy các mô hình học máy, đặc biệt như Random Forest và XGBoost, có hiệu quả vượt trội so với mô hình hồi quy tuyến tính truyền thống trong việc nâng cao độ chính xác dự báo. Các chỉ tiêu đánh giá như RMSE, MAE và  $R^2$  đều cho thấy mức cải thiện đáng kể, phản ánh khả năng của các thuật toán này trong việc xử lý dữ liệu phi tuyến và các mối quan hệ phức tạp giữa các biến đầu vào. Nghiên cứu cũng khẳng định vai trò quan trọng của các yếu tố truyền thống như quy mô công trình, chi phí vật liệu, chi phí nhân công và thời gian thi công trong việc quyết định tổng mức đầu tư. Đồng thời, việc ứng dụng BIM đã được chứng minh là có tác động tích cực trong việc giảm chi phí và nâng cao hiệu quả quản lý dự án, góp phần tối ưu hóa quá trình đầu tư xây dựng.

Ngành xây dựng đang đẩy mạnh chuyển đổi số, việc tích hợp học máy với các công cụ quản lý hiện đại như BIM và cơ sở dữ liệu lớn sẽ tạo ra một hệ thống dự báo chi phí thông minh, có khả năng hỗ trợ ra quyết định nhanh chóng và chính xác hơn. Điều này đặc biệt quan trọng đối với các dự án quy mô lớn, có tính phức tạp cao và chịu ảnh hưởng mạnh từ biến động thị trường. Bài báo cũng chỉ ra một số hạn chế nhất định, bao gồm sự thiếu hụt dữ liệu chất lượng cao, tính phân tán của dữ liệu trong ngành xây dựng, cũng như yêu cầu cao về năng lực công nghệ và nhân lực khi triển khai các mô hình học máy. Ngoài ra, một số mô hình học máy có tính chất "hộp đen" gây khó khăn trong việc diễn giải kết quả, từ đó ảnh hưởng đến khả năng áp dụng trong thực tiễn quản lý.

Như vậy, việc áp dụng học máy trong dự báo tổng mức đầu tư không chỉ nâng cao độ chính xác mà còn mở ra hướng tiếp cận mới trong quản lý chi phí xây dựng, góp phần nâng cao hiệu quả đầu tư và thúc đẩy sự phát triển bền vững của ngành xây dựng Việt Nam trong kỷ nguyên số. □

#### Tài liệu tham khảo

- [1] Bộ Xây dựng (2021), Thông tư 11/2021/TT-BXD về xác định và quản lý chi phí đầu tư xây dựng.
- [2] Chính phủ nước CHXHCN Việt Nam (2024), Nghị định 175/2024/NĐ-CP về quản lý dự án đầu tư xây dựng.
- [3] Nguyễn Thanh Bình (2023), Ứng dụng trí tuệ nhân tạo trong dự báo kinh tế - kỹ thuật.
- [4] Mai Bá Nhân và nnk (2025), Thực nghiệm ứng dụng BIM trong công tác thẩm định dự toán thiết kế tại các dự án dân dụng vốn đầu tư công. Tạp chí Vật liệu và Xây dựng, tập 15 số 1 (2025).
- [5] Nguyễn Văn Tam (2022), Ứng dụng mạng nơ-ron nhân tạo (ANN) trong dự báo chi phí xây dựng công trình dân dụng. Luận văn thạc sĩ, Trường Đại học Xây dựng Hà Nội.
- [6] Tạ Ngọc Bình (2021), Ứng dụng mô hình học máy trong dự báo chi phí xây dựng dự án đầu tư. Luận văn thạc sĩ, Đại học Giao thông Vận tải.
- [7] Ko, Y. S., & Lee, B. G. (2026), A Study on Machine Learning-Based Cost Estimation Models, Applied Sciences.
- [8] Helaly, H., et al. (2025), Machine Learning Models for Estimating Construction Costs of Conventional and Accelerated Bridge Construction Methods, The International Association for Automation and Robotics in Construction.
- [9] Abd, A. M., et al. (2024), Prediction of Highway Construction Cost using Machine Learning, Engineering, Technology and Applied Science Research (ETASR).
- [10] Alashari, M., K. El-Rayes, and H. Helaly (2023), Machine-Learning Model for Predicting Maintenance Costs of EPDM Roofing Systems, Journal of Performance of Constructed Facilities, 37 (6). American Society of Civil Engineers (ASCE). <https://doi.org/10.1061/jpcfev.cfeng-4413>.
- [11] Aprianti, E., S. Hamzah, and M. A. Abdurrahman (2021), The Analysis of Cost Estimation using Cost Significant Model on Bridge Construction in South Sulawesi, IOP Conf Ser Earth Environ Sci.