

## Đề xuất các mô hình máy học ước tính khối lượng vật tư trong giai đoạn ý tưởng dự án

Phạm Vũ Hồng Sơn<sup>1\*</sup>, Nguyễn Ngô Luân<sup>2</sup>

<sup>1</sup> Bộ môn Thi công & Quản lý Xây dựng, Trường Đại học Bách Khoa, Đại học Quốc gia-TP.HCM

<sup>2</sup> Học viên cao học Ngành Quản lý Xây dựng, Trường Đại học Bách Khoa, Đại học Quốc gia -TP.HCM

### TỪ KHÓA

Thuật toán máy học  
Weka  
Bê tông cốt thép  
Ước tính khối lượng vật tư

### TÓM TẮT

Chi phí dự án trong giai đoạn ý tưởng có vai trò quan trọng đến sự thành công của một dự án xây dựng. Các mô hình chi phí trong giai đoạn này thường chưa chi tiết hóa các chi phí vật tư, máy móc thiết bị, nhân công. Điều này đã làm các nhà quản lý dự án chưa chủ động dự trù các nguồn lực ngay từ ban đầu. Đặc biệt chi phí vật tư thường chiếm tỷ trọng lớn trong cơ cấu chi phí xây dựng các dự án dân dụng. Các nghiên cứu về ước tính khối lượng vật tư trước đây chỉ tập trung chủ yếu trong lĩnh vực: giao thông, dự án năng lượng, ... các mô hình ước tính khối lượng trong các dự án dân dụng còn hạn chế và sử dụng các phần mềm tương đối khó tiếp cận cho nhiều người trong ngành xây dựng. Bằng việc sử dụng phần mềm Weka, nghiên cứu này sẽ đề xuất các thuật toán máy học phù hợp để xây dựng mô hình ước tính khối lượng vật tư cho các dự án dân dụng kết cấu bê tông cốt thép. Kết quả dự đoán từ mô hình đề xuất sẽ được xếp hạng nhằm đề xuất các thuật toán phù hợp cho việc khai thác các mô hình bê tông, ván khuôn, cốt thép cho các cấu kiện : móng, cột, dầm và sàn.

### KEYWORDS

Machine learning algorithms  
Weka  
Reinforced concrete structures  
Materials quantity estimation

### ABSTRACT

Conceptual cost models play an important role in the success of a construction project. In this stage, the cost models often do not detail of materials, equipment and personel. This has prevented project managers from proactively planning their resources in the early of project. In particular, the cost of materials often take up a large proportion in the cost structure of civil projects. The previous studies on estimating quantities of materials only focused mainly in the fields of : traffic, plant projects, ... the models for estimating quantities in civil projects were limited and used the software that is relatively inaccessible to stakeholders in the construction industry. By using Weka software, this study will propose suitable machine learning algorithms to build a model to estimate the quantity of materials for civil projects of reinforced concrete structures. The prediction results from the proposed model will be ranked in order to propose suitable algorithms for exploiting concrete, formwork, and reinforcement models for components : foundation, column, beam and floor.

### 1. Giới thiệu

Trong giai đoạn lên ý tưởng của một dự án, các nhà phát triển dự án luôn cần những thông tin rõ ràng về chi phí xây dựng của dự án, nhưng trong khi các thông tin về bản vẽ thiết kế, đặc điểm kỹ thuật, dự toán sơ bộ thì rất hạn chế [1]. Vì vậy có thể cho rằng, việc có được thông tin về chi phí xây dựng ngay trong giai đoạn ý tưởng, sẽ hỗ trợ người phát triển dự án có những quyết định đúng đắn góp phần đạt được những mục tiêu của dự án. Nhiều mô hình đề xuất về chi phí sơ bộ đã được thực hiện, các chi phí đã được mô hình hóa với các cấp độ khác nhau về tính chính xác, đầy đủ và sự phù hợp. Các cấp độ này có thể được cải thiện bằng cách mô hình hóa các khối lượng sơ bộ - khối lượng vật tư, từ đó sẽ giúp mức độ dự báo chính xác hơn về ước tính chi phí sơ bộ trong mô hình chi phí. Mặt khác, các mô hình chi phí chủ yếu ước tính chi phí tổng hợp chưa xét đến

các yếu tố ảnh hưởng khác. Trong khi thị trường vật liệu xây dựng luôn biến động không ngừng, nếu không xét đến yếu tố vật tư thì sẽ không phản ánh chính xác về chi phí tổng hợp [2]. Hơn nữa, theo một số nghiên cứu cho thấy chi phí vật tư chiếm khoảng 42 %, một số trường hợp chiếm hơn 50 % trên tổng chi phí xây dựng tùy thuộc vào loại dự án, biện pháp thi công và phạm vi công việc [3]. Vì vậy, khối lượng vật tư và chất lượng vật tư là những yếu tố ảnh hưởng nhất đến chi phí xây dựng. [3]

Để thực hiện nghiên cứu này, đã có nhiều nghiên cứu về mô hình dựa trên AI và máy học (ML) được công bố. Máy học là một lĩnh vực phát triển của trí tuệ nhân tạo được sử dụng cho mô hình hóa các dữ liệu. Có nhiều mô hình dự báo dựa trên các dữ liệu lịch sử. Trong các mô hình máy học khác nhau, thì phổ biến là mạng nơ-ron nhân tạo (ANNs), máy véc-tơ hỗ trợ (SVMs), hồi quy đa tuyến tính [4].

\*Liên hệ tác giả: pvhson@hcmut.edu.vn

Nhận ngày 28/05/2022, sửa xong ngày 23/06/2022, chấp nhận đăng 08/09/2022

Link DOI: <https://doi.org/10.54772/jomc.05.2022.350>

Mô hình khối lượng sơ bộ trong các tòa nhà dân dụng đã được Yeh (1998) sử dụng mạng nơ-ron truyền ngược kết hợp với hồi quy thống kê để ước tính trọng lượng cốt thép trong dầm và cột, chưa ước tính khối lượng bê tông và ván khuôn cho hai cấu kiện này [5]. Các biến được dự đoán dùng trong nghiên cứu này là : số lượng tầng, bố cục lưới (grid layout), tổng chiều cao công trình, tải trọng động và tĩnh, hệ số vùng địa chấn và cường độ nén của bê tông [5]. Bakhoun và cộng sự (1998) đã sử dụng mạng nơ-ron nhân tạo để ước tính khối lượng bê tông cho một công trình về cầu ở Ai Cập [6]. Idowu và cộng sự (2020) đã sử dụng mô hình hồi quy véc-tơ hỗ trợ dựa trên nguyên lý chọn mẫu có hoàn lại (Bootstrapped Support Vector Regression Models). Nghiên cứu này cung cấp một khoảng ước tính về khối lượng (range quantities) về bê tông, thép, ván khuôn cho các cấu kiện kết cấu : móng, cột, dầm, sàn. Khoảng ước tính về khối lượng thép, ván khuôn ở mức tương đối so với thực tế, vì vậy cần nâng mức này lên cao hơn. [3]

Trong nghiên cứu này, các thuật toán máy học được đề xuất dự trên yếu tố : những phương pháp phổ biến được sử dụng trong ngành xây dựng. Nghiên cứu sẽ xem xét các thuật toán như : Mạng nơ-ron thần kinh nhân tạo (ANN), Giải thuật k-lần cận gần nhất (KNN), Máy hỗ trợ véc-tơ hồi quy (SVR) và kỹ thuật kết hợp các thuật toán máy học (Ensemble methods). Các phương pháp này đã được trình bày trong các nghiên cứu tương tự ngành xây dựng ( [7] , [8], [9] [10],). Cuối cùng, hiệu suất của các mô hình sẽ được đánh giá bởi các chỉ số như : R, RMSE, MAE, MAPE, SI. Trong phần tiếp theo sẽ mô tả các kỹ thuật máy học được xem xét, các phương pháp nghiên cứu, kết quả và thảo luận.

## 2. Phương pháp nghiên cứu

### 2.1. Các thuật toán áp dụng xây dựng mô hình

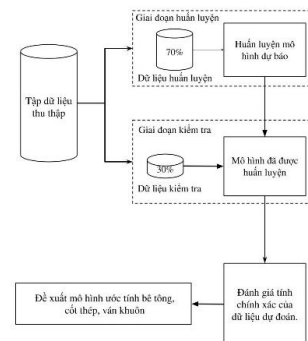
#### Mô hình dự báo dựa trên trí thông minh nhân tạo

Để xây dựng một mô hình dựa trên trí thông minh nhân tạo hiệu quả đòi hỏi cần một tiêu chuẩn cho việc xây dựng mô hình khoa học. Trong khoa học dữ liệu, mô hình tiêu biểu được lựa chọn đó là khám phá kiến thức trong cơ sở dữ liệu (KDD – Knowledge discovery in database) [11]. Các mô hình dựa trên trí thông minh nhân tạo thường được chia thành : mô hình độc lập và mô hình kết hợp. Mô hình độc lập sẽ dùng duy nhất một thuật toán, mô hình kết hợp sẽ kết hợp nhiều mô hình ước tính để ước tính kết quả đầu ra.

#### Phần mềm máy học

Các phần mềm học máy được đánh giá là mang lại độ chính xác cho các mô hình dự báo và khả năng đáp ứng các nhu cầu khác nhau của người sử dụng. Bên cạnh đó, các phần mềm này còn cung cấp giao diện thân thiện người dùng và mang tính ứng dụng cao. Các phần mềm được sử dụng phổ biến như : RapidMiner Studio, IBM SPSS Modeler, IBM SPSS Statistic, và Weka. Trong đó phần mềm Weka là công cụ được xem là dễ sử dụng, nó được viết trên nền tảng Java,

được phát triển bởi Trường Đại học Waikato, New Zealand. Weka hỗ trợ nhiều mục đích khai phá dữ liệu tiêu chuẩn như : xử lý dữ liệu, phân cụm, phân loại, hồi quy, trực quan hóa dữ liệu, ... [12]



Hình 1. Lược đồ nghiên cứu.



Hình 2. Giao diện phần mềm Weka.

#### Single Model/Các mô hình độc lập

Trong nghiên cứu này, các mô hình ước tính độc lập sẽ gồm các kỹ thuật dựa trên trí thông minh nhân tạo sẵn có. Một vài mô hình phổ biến được trình bày sau đây nhằm khám phá thêm sự hiểu biết về các mô hình dự báo độc lập như : ANNs, AANNs, SVRs, KNNs.

#### Mạng thần kinh nhân tạo và Mạng thần kinh nhân tạo bổ sung

Mô hình mạng nơ-ron nhân tạo AANNs là một thuật toán được xây dựng dựa trên sự mô phỏng mạng nơ-ron (tế bào thần kinh) của não bộ con người. Mạng truyền thẳng nhiều lớp là mạng cho kết quả dự báo tốt. Mạng truyền thẳng nhiều lớp có cách thức hoạt động như sau : các nơ-ron tại lớp đầu vào nhận tín hiệu đầu vào xử lý (tính tổng trọng số, sau đó gửi tới hàm truyền) rồi cho ra kết quả (là kết quả của hàm truyền); kết quả này sẽ được truyền tới các nơ-ron thuộc lớp ẩn thứ nhất; tại đây các nơ-ron tiếp nhận như là tín hiệu đầu vào, xử lý và gửi kết quả đến lớp ẩn tiếp theo; ... quá trình tiếp tục cho đến khi các nơ-ron thuộc lớp ra cho kết quả [13]. Việc huấn luyện sẽ được giám sát cho mọi véc-tơ đầu vào – đầu ra mong muốn, một trọng số thích nghi giảm thiểu hàm lỗi được tính toán bởi mạng dựa vào sự đo lường khác biệt giữa đầu ra dự báo và kết quả thực tế [14].

Mô hình mạng nơ-ron nhân tạo bổ sung (Additive Artificial Neural Networks –AANNs) là một siêu mô hình (meta-model) có thể

cải thiện hiệu quả của mô hình ANNs cổ điển [15]. Việc thực hiện mô hình mạng nơ-ron nhân tạo bổ sung đã được trình bày [16].

**Máy hồi quy hỗ trợ**

Máy véc-tơ hỗ trợ được phát triển bởi Vapnik vào năm 1995 dựa trên lý thuyết thống kê học và nguyên tắc tối thiểu hóa rủi ro [17]. SVMs ngày càng được sử dụng nhiều hơn trong việc giải quyết các vấn đề phi tuyến tính ngay cả với dữ liệu đưa vào huấn luyện có khối lượng nhỏ [3]. Với các vấn đề hồi quy phi tuyến, dữ liệu sẽ được chuyển đổi bằng cách sử dụng một hàm kernel phi tuyến để ánh xạ vào không gian chiều cao. Các hyperplanes tuyến tính sẽ được xây dựng thêm để hồi quy tuyến tính các dữ liệu được chuyển đổi trong không gian chiều cao. Các hàm kernel sẽ ngầm tính bên trong các biến dự đoán trong không gian đó. Vì vậy, hiệu suất của mô hình này phụ thuộc vào lựa chọn tham số kernel. Mục tiêu của máy véc-tơ hỗ trợ hồi quy là tìm hàm hồi quy dựa trên một tập dữ liệu đầu vào dùng để dự đoán giá trị mong muốn (đầu ra) [3].

Idowu và cộng sự đã sử dụng SVM để mô hình hóa một thành tố của chi phí – đó là khối lượng trong giai đoạn lên ý tưởng của dự án. Hiệu suất của hồi quy vectơ hỗ trợ tốt hơn nhưng còn kém so với giải thuật KNN [3]...

**Giải thuật lân cận k-gần nhất (k-Nearest Neighbors- KNN)**

Giải thuật lân cận k-gần nhất là một trong những giải thuật đơn giản và được sử dụng rộng rãi dựa trên dữ liệu lịch sử để xác định điểm lân cận gần nhất của một điểm dữ liệu được cho. Trong thuật toán này, để tính toán một kết quả dự đoán, các thực thể huấn luyện (training instances) nào gần với quan sát mới nhất sẽ được sử dụng. Vì vậy, sự chính xác của kết quả dự đoán phụ thuộc nhiều vào giá trị của k [18]. Các tập dữ liệu có kích thước càng lớn thì k có giá trị càng lớn.

**Ensemble Model/Các mô hình kết hợp**

Bằng việc kết hợp các mô hình dự báo với nhau để cho ra kết quả dự báo được chính xác hơn, các mô hình kết hợp đang thu hút sự chú ý trong cộng đồng sử dụng các thuật toán máy học. Sự kết hợp này giúp mô hình dự báo vượt trội hơn so với các mô hình độc lập [19]. Các mô hình kết hợp được đề xuất sau đây giúp khai phá trong việc ứng dụng ước tính trong nghiên cứu này. Mô hình kết hợp các thuật toán khác nhau như: ANN-SVRs, ANN-KNNs, ANN-SVRs-LR-CART, ANN-KNNs-LR-CART.

**Evaluation Parameters/Các chỉ số đánh giá mô hình**

Các chỉ số như R, RMSE, MAE, MAPE và SI được sử dụng để đánh giá tính chính xác của các mô hình ước tính được đề xuất [14]. Các công thức từ (1) đến (6) của các chỉ số sẽ được thể hiện như sau:

$$R = \frac{n \sum y.y' - (\sum y).(\sum y')}{\sqrt{(n \sum y^2 - (\sum y)^2)(n \sum y'^2 - (\sum y')^2)}} \quad (1)$$

trong đó y' là giá trị ước tính, y là giá trị thực tế, n là số dữ liệu mẫu.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - y'| \quad (2)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y - y'}{y} \right| \quad (3)$$

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - y')^2} \quad (4)$$

$$MaxAE = \max(|y - y'|) \quad (5)$$

$$SI = \frac{1}{m} \sum_{i=1}^m \left( \frac{P_i - P_{min,i}}{P_{max,i} - P_{min,i}} \right) \quad (6)$$

trong đó m là số lượng đo lường hiệu suất và P<sub>i</sub> = i<sup>th</sup> đo lường hiệu suất. Chỉ số SI sẽ nằm trong khoảng từ 0 đến 1. Chỉ số SI càng tiến về 0 thì mô hình ước tính càng cho kết quả ước tính chính xác.

**2.2. Nguồn dữ liệu và Phát triển mô hình**

**2.2.1 Nguồn dữ liệu**

Dữ liệu được thu thập từ 80 dự án (48 dự án dân dụng, 32 dự án thương mại) từ 7 công ty thiết kế kết cấu có kinh nghiệm từ 6-17 năm trong lĩnh vực này tại Nigeria. Các chi tiết thiết kế bản vẽ kỹ thuật các dự án đều đã được phê duyệt và được cung cấp dưới dạng bảng tính kết quả kết cấu [3].

**2.2.2 Cài đặt và thực thi các mô hình đề xuất**

Bảng 1 trình bày tóm tắt về các thuộc tính cho các thuật toán máy học được đề xuất sử dụng trong nghiên cứu này. Các mô hình này bao gồm: SVR-PL, SVR-RBF, ANN – 1 lớp ẩn, ANN – 2 lớp ẩn, AANN – 1 lớp ẩn, AANN – 2 lớp ẩn, KNN – k=2, KNN – k=3, ANN-SVR(PL), ANN-SVR(RBF), ANN-SVR(RBF) - LR – CART, ANN-SVR(PL) - LR – CART, ANN-KNN(k=2) - LR – CART, ANN-KNN(k=3) - LR – CART. Các mô hình này sẽ được thực hiện trên nền tảng phần mềm Weka.

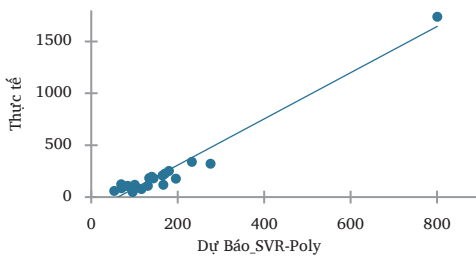
**Bảng 1.** Các thuộc tính cho mô hình.

Biến phản hồi	Đơn vị	Các biến dự đoán
Mong_BeTong_(FNDN_CONC)	m <sup>3</sup>	GFL + GSR + SBC + NF
Dam_BeTong_(Beam_CONC)	m <sup>3</sup>	GFL + BF + NF
San_BeTong_(Slab_CONC)	m <sup>3</sup>	GFL + BF + NF
Cot_BeTong_(Col_CONC)	m <sup>3</sup>	GFL + BF + NF
Mong_CotThep_(FNDN_RFT)	kg	NF + GFL + GSR + SBC
Dam_CotThep_(Beam_RFT)	kg	GFL + BF + NF
San_CotThep_(Slab_RFT)	kg	GFL + BF + NF
Cot_CotThep_(Col_RFT)	kg	BF + GFL + NF
Mong_VanKhuon_(FNDN_FMK)	m <sup>2</sup>	GSR + GFL + SBC + Qk
Dam_VanKhuon_(Beam_FMK)	m <sup>2</sup>	GFL + BF + NF
San_VanKhuon_(Slab_FMK)	m <sup>2</sup>	GFL + BF + NF
Cot_VanKhuon_(Col_FMK)	m <sup>2</sup>	BF + GFL + Qk
Ghi chú		
GFL: Tổng tải trọng sàn, GFA: Tổng diện tích sàn, Qk: tải trọng động		
GSR: Tổng ứng xử đất, SBC: Khả năng chịu lực của đất		
BF: Diện tích xây dựng, NF: Chiều cao, số tầng		

Trong các mô hình máy học, nhằm giúp các mô hình được khái quát hóa tốt tránh hiện tượng học quá và học thiếu thì dữ liệu ban đầu được chia làm 2 phần: tập dữ liệu huấn luyện và dữ liệu kiểm tra. Để kết quả ước tính cho kết quả chính xác hơn, Kohavi (1995) [20] đề xuất sử dụng xác thực chéo 10-lớp cho mô hình. Trong nghiên cứu này, các dữ liệu thu thập được sẽ được chia ngẫu nhiên thành 2 tập dữ liệu bằng phần mềm Weka trong đó: 70 % tập dữ liệu huấn luyện (56 dự án) và 30 % tập dữ liệu kiểm tra (24 dự án). Tập dữ liệu huấn luyện sẽ được sử dụng để tạo các mô hình độc lập (ANN, AANNs, SVRs, KNNs) và các mô hình kết hợp (ANN-SVRs, ANN-KNNs, ANN-SVRs-LR-CART, ANN-KNNs-LR-CART) trong khi đó tập dữ liệu kiểm tra được xem như dữ liệu “chưa biết” dùng để đánh giá hiệu quả của các mô hình được đề xuất. Luật học có giám sát và xác thực chéo 10-lớp (10-folds) sẽ được thực hiện trong tập dữ liệu huấn luyện. Các mô hình được đề xuất sẽ được xếp hạng thông qua chỉ số SI. Các mô hình có chỉ số SI thấp nhất sẽ là mô hình cho kết quả dự đoán chính xác nhất trong các mô hình đề xuất.

### 3. Kết quả và thảo luận

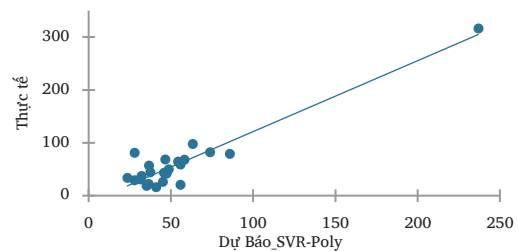
Các mô hình máy học được đề xuất trong mô hình như: các mô hình độc lập (ANN, AANNs, SVRs, KNNs) và các mô hình kết hợp



Hình 3. Biểu đồ phân tán kết quả dự đoán mô hình SVR-Poly khối lượng bê tông sàn.

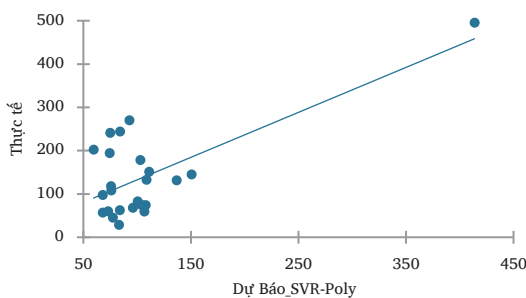
(ANN-SVRs, ANN-KNNs, ANN-SVRs-LR-CART, ANN-KNNs-LR-CART) được sử dụng ước tính khối lượng trong giai đoạn lên ý tưởng bằng dữ liệu thu thập được từ 80 dự án. Sau khi thực hiện, tính chính xác của mô hình sẽ được thể hiện trong Bảng 2. Thông qua các chỉ số R, RMSE, MAE, MAPE trong hai giai đoạn: huấn luyện và kiểm tra.

Trong các mô hình được đề xuất, có thể thấy các mô hình độc lập và mô hình kết hợp đều cho kết quả dự đoán khá tốt. Với các mô hình ước tính khối lượng bê tông cho các cấu kiện cột, dầm và sàn thì mô hình SVR-Poly cho kết quả ước tính chính xác nhất. Trong đó mô hình SVR-Poly ước tính khối lượng bê tông cột có chỉ số SI=0,015 thấp nhất trong các mô hình được đề xuất, MAPE = 44,76 %; RMSE=26,450 m<sup>3</sup>, MAE=19,123 m<sup>3</sup> trong giai đoạn huấn luyện và trong giai đoạn kiểm tra các chỉ số này lần lượt là: 39,93 m<sup>3</sup>; 24,726 m<sup>3</sup>; 16,908 m<sup>3</sup>. Hình 3 cho thấy mức độ tập trung của dữ liệu là tốt khi giá trị ước tính và thực tế có sự tương quan với nhau rất cao. Với mô hình SVR-Poly ước tính bê tông cấu kiện sàn, mô hình này có chỉ số SI thấp nhất là 0,0195. Trong giai đoạn kiểm tra hiệu quả của mô hình bằng các chỉ số MAPE=26,848 %, RMSE=196,55 m<sup>3</sup>; MAE=74,758 m<sup>3</sup> cho kết quả khá tốt. Hình 4 biểu diễn kết quả giữa dự đoán – thực tế của mô hình cho thấy độ tập trung của dữ liệu khá cao.

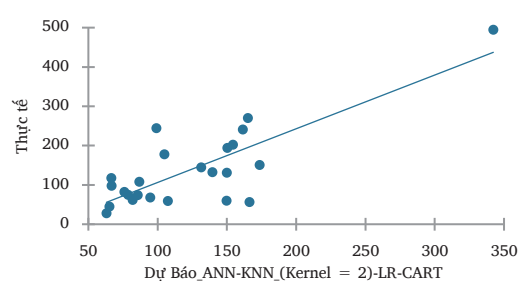


Hình 4. Biểu đồ phân tán kết quả dự đoán mô hình SVR-Poly khối lượng bê tông cột.

Trong mô hình ước tính khối lượng bê tông móng các mô hình kết hợp lại phát huy hết các điểm mạnh của chúng khi 4/5 thuật toán đề xuất tốt nhất là các mô hình kết hợp. Trong đó mô hình ANN-KNN(k=2) - LR-CART có chỉ số SI thấp nhất là 0,027. Chỉ số MAPE = 56,49 %; RMSE = 55,12 m<sup>3</sup>; MAE = 39,83 m<sup>3</sup> trong giai đoạn huấn luyện, trong giai đoạn kiểm tra thì mô hình này có các chỉ số MAPE, RMSE, MAE lần lượt là 45,158 %; 65,110 m<sup>3</sup>; 49,285 m<sup>3</sup>. Mặc dù trong giai đoạn kiểm tra thì chỉ số MAPE cho kết quả tốt nhưng các chỉ số còn lại cao hơn so với trong giai đoạn huấn luyện. Hình 6 đã cho thấy điều này, khi biểu đồ phân tán mô tả giá trị thực tế và dự đoán có xu hướng phân tán.



Hình 5. Biểu đồ phân tán kết quả dự đoán mô hình SVR-Poly khối lượng bê tông dầm.



Hình 6. Biểu đồ phân tán kết quả dự đoán mô hình ANN-KNN\_ (Kernel = 2)-LR-CART khối lượng bê tông móng.

**Bảng 2.** Độ chính xác dự đoán của mô hình cho ước tính khối lượng vật tư (bê tông).

Các chỉ số đánh giá sự chính xác trong giai đoạn huấn luyện Tập dữ liệu huấn luyện (N = 56 dự án, 70 %)					Các chỉ số đánh giá sự chính xác trong giai đoạn kiểm tra Tập dữ liệu kiểm tra (N = 24 dự án, 30%)			
Model	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)
<b>DÀM</b>								
SVR-RBF	0,717	53,244	38,027	45,882%	0,658	91,328	65,440	56,083%
SVR-Poly	0,856	35,255	26,374	31,797%	0,703	78,378	57,938	46,323%
ANN-1HL	0,788	42,266	27,806	28,882%	0,682	93,055	64,254	39,392%
ANN-2HL	0,802	40,258	26,989	28,053%	0,631	89,880	65,533	48,088%
AANN-1HL	0,774	43,792	29,294	29,835%	0,687	82,528	60,527	43,731%
AANN-2HL	0,808	39,717	27,886	29,689%	0,630	84,934	64,434	51,697%
KNN-K = 2	0,793	41,650	28,047	29,387%	0,524	95,199	71,467	53,579%
KNN-K = 3	0,835	38,096	25,456	27,320%	0,452	98,722	74,610	57,439%
ANN-SVR_(PoLy)	0,838	36,996	25,248	28,236%	0,694	85,023	59,776	40,676%
ANN-SVR_(RBF)	0,784	43,337	28,745	33,015%	0,677	88,436	63,145	45,163%
ANN-SVR_(RBF)-LR-CART	0,847	36,939	25,291	29,759%	0,674	86,300	61,253	43,065%
ANN-SVR_(Poly)-LR-CART	0,859	34,478	24,344	28,089%	0,685	84,780	59,873	40,980%
ANN-KNN_(Kernel = 2)-LR-CART	0,857	34,910	24,451	27,254%	0,645	88,956	63,393	42,975%
ANN-KNN_(Kernel = 3)-LR-CART	0,863	34,289	23,727	26,652%	0,634	89,685	64,041	43,759%
<b>SÀN</b>								
Model	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)
SVR-RBF	0,777	96,711	65,197	49,654%	0,962	292,944	108,317	44,023%
SVR-Poly	0,905	55,488	35,731	26,149%	0,980	196,550	74,758	26,848%
ANN-1HL	0,807	82,727	49,146	33,178%	0,984	196,622	78,934	28,686%
ANN-2HL	0,836	71,858	50,388	37,566%	0,983	210,884	87,062	30,762%
AANN-1HL	0,797	85,347	48,917	31,482%	0,976	205,472	92,709	38,657%
AANN-2HL	0,841	71,173	49,283	36,608%	0,983	206,195	82,738	30,547%
KNN-K = 2	0,860	70,202	41,267	29,392%	0,951	236,599	93,264	32,500%
KNN-K = 3	0,883	66,165	38,867	27,999%	0,928	248,010	94,331	34,324%
ANN-SVR_(PoLy)	0,879	63,141	41,330	28,790%	0,982	196,485	76,846	27,767%
ANN-SVR_(RBF)	0,805	75,626	50,303	37,509%	0,981	243,725	86,828	29,124%
ANN-SVR_(RBF)-LR-CART	0,894	60,980	43,142	34,236%	0,976	236,968	85,778	27,527%
ANN-SVR_(Poly)-LR-CART	0,905	54,318	39,059	29,462%	0,978	213,477	81,332	27,433%
ANN-KNN_(Kernel = 2)-LR-CART	0,911	54,553	37,764	29,168%	0,977	223,191	85,919	28,823%
ANN-KNN_(Kernel = 3)-LR-CART	0,910	54,576	37,915	28,756%	0,973	226,109	85,948	28,970%
<b>CỘT</b>								
Model	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)
SVR-RBF	0,706	30,073	21,072	60,730%	0,94	47,682	26,557	57,846%
SVR-Poly	0,668	26,450	19,123	44,756%	0,94	24,726	16,908	39,935%
ANN-1HL	0,679	25,661	18,573	44,700%	0,85	32,560	20,412	41,444%
ANN-2HL	0,656	26,482	19,647	48,265%	0,82	34,938	21,134	39,958%
AANN-1HL	0,681	25,589	19,019	44,481%	0,84	32,113	21,688	46,702%
AANN-2HL	0,676	25,696	19,187	48,128%	0,84	33,338	21,376	45,477%
KNN-K = 2	0,697	25,087	19,167	44,433%	0,71	43,136	25,363	49,405%
KNN-K = 3	0,739	23,919	17,495	42,687%	0,72	43,569	24,253	44,930%
ANN-SVR_(PoLy)	0,710	24,606	18,325	43,689%	0,92	27,560	17,633	39,041%
ANN-SVR_(RBF)	0,711	26,248	18,247	49,777%	0,88	38,096	21,279	45,927%
ANN-SVR_(RBF)-LR-CART	0,765	23,599	17,074	45,884%	0,87	37,018	20,337	42,863%

ANN-SVR_(Poly)-LR-CART	0,760	22,713	16,797	42,556%	0,89	31,695	18,767	39,944%
ANN-KNN_ (Kernel = 2)-LR-CART	0,767	22,655	16,809	42,077%	0,83	36,568	21,532	43,335%
ANN-KNN_ (Kernel = 3)-LR-CART	0,773	22,487	16,586	41,884%	0,84	36,606	20,961	41,691%
<b>MÓNG</b>								
Model	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)	R	RMSE (m <sup>3</sup> )	MAE (m <sup>3</sup> )	MAPE (%)
SVR-RBF	0,686	60,442	44,962	57,167%	0,729	93,943	67,524	66,416%
SVR-Poly	0,636	59,430	44,150	59,291%	0,799	73,219	57,739	49,432%
ANN-1HL	0,545	65,817	47,084	65,630%	0,747	78,342	59,715	45,055%
ANN-2HL	0,527	76,006	60,374	74,883%	0,724	87,158	61,381	42,607%
AANN-1HL	0,408	79,277	54,599	70,884%	0,762	72,867	55,567	44,387%
AANN-2HL	0,489	77,008	59,737	74,104%	0,742	81,325	57,919	42,716%
KNN-K=2	0,321	76,662	58,416	64,647%	0,643	77,930	57,224	59,807%
KNN-K=3	0,398	69,504	52,547	65,079%	0,609	79,778	56,525	56,290%
ANN-SVR_(PoLy)	0,664	53,930	37,916	57,976%	0,785	73,877	55,006	43,035%
ANN-SVR_(RBF)	0,604	55,220	39,708	57,303%	0,755	79,881	57,781	48,349%
ANN-SVR_(RBF)-LR-CART	0,691	50,410	36,294	54,726%	0,774	72,438	55,474	49,380%
ANN-SVR_(Poly)-LR-CART	0,698	50,494	35,424	55,524%	0,788	69,003	52,875	45,172%
ANN-KNN_ (Kernel = 2)-LR-CART	0,625	55,123	39,821	56,449%	0,803	65,110	49,285	45,158%
ANN-KNN_ (Kernel = 3)-LR-CART	0,634	54,657	38,716	56,879%	0,799	66,626	49,694	44,511%

**Bảng 3.** Xếp hạng các mô hình đề xuất cho ước lượng khối lượng (bê tông).

Cấu kiện	Mô hình	SI
Móng	ANN-KNN_ (Kernel = 2)-LR-CART	0,027
	ANN-KNN_ (Kernel = 3)-LR-CART	0,044
	ANN-SVR_(Poly)-LR-CART	0,129
	ANN-SVR_(PoLy)	0,183
	AANN-1HL	0,225
Cột	SVR-Poly	0,015
	ANN-SVR_(PoLy)	0,114
	ANN-SVR_(Poly)-LR-CART	0,272
	ANN-1HL	0,401
	ANN-2HL	0,483
Dầm	SVR-Poly	0,096
	AANN-1HL	0,230
	ANN-SVR_(PoLy)	0,238
	ANN-SVR_(Poly)-LR-CART	0,246
	ANN-SVR_(RBF)-LR-CART	0,348
Sàn	SVR-Poly	0,020
	ANN-SVR_(PoLy)	0,038
	ANN-1HL	0,058
	ANN-SVR_(Poly)-LR-CART	0,147
	AANN-2HL	0,156

#### 4. Kết luận và đề xuất

Với các kết quả đạt được từ các mô hình máy học được đề xuất, phần lớn các mô hình được lựa chọn cho thấy việc ước tính khối lượng sơ bộ cho kết quả khá tốt, cho dù trong giai đoạn ý tưởng dự án, các thông tin về dự án điều khiển khuyết. Nghiên cứu này hỗ trợ việc lập kế hoạch dự trù các nguồn lực cần thiết trong giai đoạn sớm của dự án khi mà trước đây các chi phí xây dựng chưa được chi tiết hóa (đặc biệt chi phí vật tư). Thông qua các kết quả có được, nghiên cứu cho thấy sự thân thiện về giao diện người dùng của phần mềm Weka cho nhiều đối tượng trong ngành xây dựng với mục đích ước tính các khối lượng sơ bộ ngay từ giai đoạn ý tưởng của dự án.

Nghiên cứu này mở ra một số hướng nghiên cứu có thể thực hiện trong tương lai như: sử dụng một tập dữ liệu cho các dự án dân dụng cao tầng, các loại dự án khác; Áp dụng các mô hình đề xuất cho công tác hoàn thiện, MEP...; Khám phá thêm các thuật toán máy học độc lập, kết hợp hay các mô hình lai khác (hybrid model); Xét thêm các biến đưa vào mô hình để kết quả dự đoán được cải thiện hơn.

**Lời cảm ơn:** Nghiên cứu được hỗ trợ bởi Trường Đại học Bách Khoa - ĐHQG-HCM.

#### Tài liệu tham khảo

- [1] M. Badawy, "A hybrid approach for a cost estimate of residential buildings in Egypt at the early stage," *Asian Journal of Civil Engineering*, 2020. z
- [2] H.-S. L. M. P. D.-Y. H. a. J. A. Bo-Sik Son, "Quantity Based Active Schematic Estimating (Q-BASE) Model," *KSCE Journal of Civil Engineering*, 2013.
- [3] O. S. I. a. K. C. Lam, "Conceptual Quantities Estimation Using Bootstrapped

- Conceptual Quantities Estimation Using Bootstrapped," *Journal of Construction Engineering and Management*, 2020.
- [4] S. A. ., M. N. W. S. A. S. S. a. G. H. C. Sparsh Sharma 1, "A Survey on Applications of Artificial Intelligence for Pre-Parametric Project Cost and Soil Shear-Strength Estimation in Construction and Geotechnical Engineering," *MDPI*, 2021.
- [5] I.-C. Yeh, "Quantity Estimating Of Building With Logarithm-Neuron Networks," *Journal of Construction Engineering and Management*, 1998.
- [6] G. S. L. M. P. D. M. E. -S. D. M. A. T. Dr. Mourad M. Bakhoun, "Estimation Of Quantities And Cost Of Prestressed Concrete Bridges Over The Nile In Egypt," *EGYPTIAN SOCIETY OF ENGINEERS*, 1998.
- [7] Attal, "Ohio Library and Information Network (OhioLINK)," Ohio, 2010. [Online]. Available: [https://etd.ohiolink.edu/apexprod/rws\\_olink/r/1501/10?clear=10&p10\\_accession\\_num=ohiou1282146503](https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=ohiou1282146503). [Accessed 5 10 2022].
- [8] A. Czarnigowska and A. Sobotka, "Estimating Construction Duration for Public Roads During the Preplanning Phase," *Journal of Engineering, Project, and Production Management*, vol. 4(1), pp. 26-35, 2014.
- [9] S. Bayram, "Duration Prediction Models for Construction Projects: In Terms of Cost or Physical Characteristics?," *KSCE Journal of Civil Engineering*, vol. 21, pp. 2049-2060, 2017.
- [10] Igor Pesko, Vladimir Mucenski, Milos Seslija, Nebojsa Radovic, Aleksandra Vujkov, Dragana Bibic and Milena Krklješ, "Estimation of Costs and Durations of Construction of Urban Roads Using ANN and SVM," *Complexity*, vol. 2017, 2017.
- [11] X. He, P. A. Rui Liu and P. F. and Chimay J. Anumba, "Data-Driven Insights on the Knowledge Gaps of Conceptual Cost Estimation Modeling," *Journal of Construction Engineering and Management*, 2021.
- [12] Waikato University, [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/>.
- [13] T. N. C. Kiên, "Viện Hàn Lâm Khoa Học Và Công Nghệ," 2017. [Online]. Available: <https://imech.ac.vn/upload/NewsImage/2021/1/12/ung-dung-mang-noron-nhan-tao-vao-bai-toan-du-bao-thuy-van.pdf>. [Accessed 10 11 2021].
- [14] Jui-Sheng Chou, Duc-Son Tran, "Forecasting Energy Consumption Time Series using Machine Learning Techniques based on Usage Patterns of Residential Householders," *Energy*, vol. 165, pp. 709-726, 2018.
- [15] J. H. Friedman, "Stochastic gradient boosting," *COMPUTATIONAL STATISTICS & DATA ANALYSIS*, 2002.
- [16] N.-T. N. ., a. A.-D. P. Ngoc-Son Truong, "Forecasting Time-Series Energy Data in Buildings Using an Additive Artificial Intelligence Model for Improving Energy Efficiency," *Hindawi*, 2021.
- [17] Vladimir Vapnik & Corinna Cortes , "Support-vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [18] B. Yu, P. Xiaolin Song, P. Feng Guan and P. Zhiming Yang, "k -Nearest Neighbor Model for Multiple-Time-Step Prediction of Short-Term Traffic Condition," *Journal of Transportation Engineering*, vol. 142, no. 6, 2016.
- [19] Murat Ayhan, Irem Dikmen, and Talat Birgonul, "Predicting the Occurrence of Construction Disputes Using Machine Learning Techniques," *Journal of Construction Engineering and Management*, vol. 147, 2021.
- [20] Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the 14th international joint conference on Artificial intelligence*, vol. 2, pp. 1137-1143, 1995.