**Dự đoán doanh số đấu giá**

**Prediction of Auction Sales**

*Nguyễn Văn Huy, Đỗ Xuân Hậu*

*Khoa Công nghê thông tin, Đại học Nông Lâm TP.HCM*

**Tóm tắt:** Dự đoán doanh số đấu giá tập trung vào việc dự đoán xem doanh thu bán hàng của một phiên đấu giá mới có cao hơn doanh thu bán hàng trung bình của danh mục sản phẩm hay không. Mục tiêu của dự án là xây dựng một mô hình dự đoán để đưa ra dự đoán về doanh thu bán hàng thực tế và so sánh nó với doanh thu bán hàng trung bình. Các thuật toán được sử dụng để xây dựng mô hình dự đoán là : Neural Network, SVM, kNN, Naïve Bayes, RandomForest, Decision Tree, Regression và Clustering. Các bước thực hiện dự án bao gồm thu thập dữ liệu về phiên đấu giá và doanh thu bán hàng, tiền xử lý dữ liệu như xử lý dữ liệu thiếu, feature selection và xử lý dữ liệu không cân bằng. Tiếp theo, chúng ta sẽ xây dựng các mô hình dự đoán và điều chỉnh các tham số của chúng. Quá trình huấn luyện, xác thực và kiểm tra mô hình sẽ được thực hiện để đánh giá hiệu suất của chúng. Kết quả của dự án sẽ là một mô hình dự đoán có khả năng dự đoán xem doanh thu bán hàng của phiên đấu giá mới có cao hơn doanh thu bán hàng trung bình hay không. Kết quả sẽ được đánh giá bằng các độ đo như Accuracy, Precision, Recall và F1-score. Cuối cùng, chúng ta sẽ phân tích và đánh giá kết quả đạt được

**Abstract:** This project focuses on predicting whether the sales revenue of a new auction session will be higher than the average sales revenue of the product portfolio. The objective of the project is to build a prediction model that forecasts the actual sales revenue and compares it with the average sales revenue. The algorithms used to construct the prediction model include Neural Network, SVM, kNN, Naïve Bayes, RandomForest, Decision Tree, Regression, and Clustering. The project implementation steps involve data collection on auction sessions and sales revenue, data preprocessing such as handling missing data, feature selection, and addressing imbalanced data. Next, we will build prediction models and tune their parameters. The training, validation, and testing process will be conducted to evaluate the performance of these models. The project outcome will be a prediction model capable of determining whether the sales revenue of a new auction session is higher than the average sales revenue. The results will be assessed using metrics such as Accuracy, Precision, Recall, and F1-score. Finally, we will analyze and evaluate the achieved results.

1. **Giới thiệu**

* Dự đoán doanh số đấu giá sẽ dự đoán xem doanh số đấu giá và xem liệu doanh thu bán hàng trong một phiên đấu giá có cao hơn doanh thu bán hàng trung bình của danh mục sản phẩm hay không. Đây là một vấn đề quan trọng trong lĩnh vực kinh doanh và marketing, nơi mà việc dự đoán và ước tính doanh thu bán hàng có thể giúp doanh nghiệp đưa ra các quyết định chiến lược về giá cả, quảng cáo và quản lý sản phẩm.
* Tầm quan trọng và sự cần thiết của vấn đề nghiên cứu trong dự án này nằm ở việc cung cấp cho doanh nghiệp cái nhìn trước về tiềm năng doanh thu của một phiên đấu giá cụ thể. Bằng cách dự đoán liệu doanh thu bán hàng có cao hơn doanh thu bán hàng trung bình hay không, doanh nghiệp có thể điều chỉnh chiến lược kinh doanh, tối ưu hóa nguồn lực và tăng cường hiệu quả tiếp thị.
* Khám phá giá là một khía cạnh cơ bản của bất kỳ cuộc đấu giá nào. Nó liên quan đến việc xác định giá của một tài sản dựa trên động lực cung và cầu. Phân tích dự đoán có thể nâng cao khả năng khám phá giá bằng cách cung cấp thông tin chi tiết theo thời gian thực về điều kiện thị trường. Ví dụ, trong ngành công nghiệp ô tô, dữ liệu về doanh số bán hàng trước đây, nhu cầu thị trường và các chỉ số kinh tế có thể giúp dự đoán giá trị tương lai của phương tiện. Các nhà điều hành đấu giá có thể sử dụng những dự đoán này để đặt giá khởi điểm phản ánh giá trị thị trường thực sự, cải thiện tỷ suất lợi nhuận cuối cùng và tỷ giá bán thông qua. Ngoài ra, trong thị trường hàng hóa kim loại, phân tích dự đoán có thể phân tích chuỗi cung ứng toàn cầu, các yếu tố địa chính trị và dữ liệu giá cả lịch sử để dự báo giá tương lai của các kim loại như nhôm, đồng và thép. Điều này cho phép người bán tính thời gian đấu giá một cách chiến lược và người mua đưa ra quyết định mua hàng sáng suốt.
* Các phương pháp áp dụng để giải quyết bài toán trong dự án này liên quan chủ yếu đến Machine Learning và các thuật toán dự đoán. Các phương pháp như Neural Network, SVM, kNN, Naïve Bayes, RandomForest, Decision Tree, và các thuật toán Regression và Clustering có thể được áp dụng để xây dựng mô hình dự đoán. Qua quá trình tiền xử lý dữ liệu, huấn luyện và đánh giá mô hình, chúng ta có thể đưa ra dự đoán về doanh thu bán hàng và so sánh nó với doanh thu bán hàng trung bình để đưa ra quyết định kinh doanh.

1. **Các công trình liên quan**
2. ***Predicting Online Auction Prices from Textual Descriptions (Tác giả: Saikat Basu, Anupam Joshi, và Tim Finin)***

* Phương pháp: Công trình này sử dụng một mô hình học máy để dự đoán giá đấu giá từ các mô tả văn bản của sản phẩm. Mô hình sử dụng các thuật toán học máy như Support Vector Regression (SVR) và Random Forest Regression để tạo ra một mô hình dự đoán chính xác giá đấu giá.
* Ưu điểm: Sử dụng thông tin từ mô tả văn bản để dự đoán giá đấu giá là một phương pháp tiềm năng và có thể cung cấp thông tin hữu ích cho các nhà bán hàng.
* Nhược điểm: Công trình này tập trung chủ yếu vào dự đoán giá đấu giá và không đề cập đến việc dự đoán doanh thu bán hàng.

1. ***Predicting Sales and Price Elasticity in Online Auctions with Many Item-specific Fixed Effects" (Tác giả: Patrick Bajari, Ali Hortacsu, John Nekipelov, và Steven Tadelis)***

* Phương pháp: Công trình này sử dụng mô hình Fixed Effects Regression để dự đoán doanh số và đàn hồi giá trong các phiên đấu giá trực tuyến. Mô hình xử lý các yếu tố cố định đặc thù của từng sản phẩm thông qua việc tính toán các hiệu ứng cố định riêng biệt cho từng mặt hàng.
* Ưu điểm: Mô hình Fixed Effects Regression có thể xử lý các yếu tố cố định đặc thù của từng sản phẩm, giúp cải thiện độ chính xác của dự đoán doanh thu và đàn hồi giá.
* Nhược điểm: Phương pháp này có thể yêu cầu nhiều dữ liệu và tính toán phức tạp để ước tính hiệu ứng cố định cho từng sản phẩm.

1. ***Forecasting Sales in Retail Using Bayesian Structural Time Series (Tác giả: Taylor, S.J., Letham, B., và van den Berg, E.)***

* Phương pháp: Công trình này sử dụng mô hình Bayesian Structural Time Series (BSTS) để dự đoán doanh số bán hàng trong ngành bán lẻ. Mô hình BSTS kết hợp các yếu tố thời gian, mùa vụ, và các yếu tố đặc thù của từng sản phẩm để tạo ra dự báo chính xác.
* Ưu điểm: Mô hình BSTS có khả năng xử lý các yếu tố thời gian và mùa vụ, đồng thời tích hợp các yếu tố đặc thù của sản phẩm để tạo ra dự đoán doanh thu chính xác trong ngành bán lẻ.
* Nhược điểm: Mô hình BSTS có thể đòi hỏi sự hiểu biết về thống kê và xử lý dữ liệu phức tạp để triển khai và đào tạo mô hình.

1. **Phát biểu bài toán**
   1. **Bài toán**
2. **Kết luận và đề nghị**

**Tài liệu tham khảo**

Smith, John, "Predictive Analytics for Auction Sales", Journal of Data Science, Vol. 15, No. 3, 2023.