**Predicting the price of a sale**

**based on customer information**

**Machine Learning**

Insper – 2023.1

Natália Queiroz Menezes Carreras

Willian Kenzo Asanuma Lee

**Introduction:**

Machine Learning has become a great way for companies to gain valuable insights into consumer behavior by making predictions about sales depending on the time of year, location of the store and personal information about their clients, such as age and gender. In addition, when gaining these valuable insights, companies will be able to create and improve strategies to optimize their gains and expenses, in all areas of their businesses. Taking this into consideration, we created a model capable of predicting how much a person is going to spend based on all the information gathered.

**Research data and analysis:**

To start the training model, we chose a dataset called *“Customer Shopping Dataset – Retail Sales Data”*[1]  which analysis the Istanbul Market. The dataset includes invoice number, customer ID, age, gender, payment method, product category, quantity, price, order date and the location of the shopping mall from 10 different malls between 2021 and 2023.

Gráfico, Gráfico de pizza

Descrição gerada automaticamenteGráfico, Gráfico de pizza

Descrição gerada automaticamente We started by analyzing each column separately and plotting graphs to analyze the information better. All the graphs are shown below. By analyzing the category of the items purchased, it is possible to see that 34.68% are clothing, followed by 15.18% that are cosmetics and then 14.86% are Food and Beverages. After, we started analyzing the gender information about the clients, being able to conclude that most clients are *Female,* reaching almost 60% of all customers. When analyzing the payment method, we were able to notice that *cash* is the most used method by customers (44.7%), then *credit card* (35.1%) and the least used is *debit card* (20.2%) as seen in the pie chart below.

Figure 1: Category of the items purchased.

After analyzing the information about the customers, we began analyzing locations and date of the purchase. Taking into consideration all the purchases, we were able to figure out in which months most of the sales are made and in which month the sales are the least. Furthermore, we were also able to get the amount gained in each month off sales. By doing so, we were able to conclude that the most profitable months are *October, November* and *December*.

Figure 2: Payment method.

Aplicativo

Descrição gerada automaticamente

Figure 3: Purchases made by month (between 2021 and 2023)

Interface gráfica do usuário, Gráfico, Aplicativo

Descrição gerada automaticamente

Figure 4: Amount of money gained on sales.

Aplicativo

Descrição gerada automaticamente

**Conclusion:**

**Bibliography:**

[1] MEHMET TAHIR ASLAN. **Customer Shopping Dataset - Retail Sales Data**. Disponível em: <https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset> . Acesso em: 25 maio. 2023.