UNIVERSIDAD ANÁHUAC QUERÉTARO

MIDTERM EXAM

SUBJECT: Aprendizaje de Máquina
CAREER: **Ingeniería Mecatrónica.**
ID: SIS-4309
SEMESTER: 6°
PROFESSOR: Jonathan Domínguez Aldana
DATE: 03/03/2020
Student Name: _____
Student ID: _____

INSTRUCTIONS

1. SCORING INSTRUCTIONS: The exam comprises 100 points distributed in different sections. The value of each section and their corresponding instructions are indicated at the start of each one of them.

2. ALLOWED RESOURCES: The usage of any gadgets and communication media through social networks, phone, messenger, or mail are not allowed during the exam. This activity will be considered as an act of academic dishonesty.

3. DURATION: The maximum time to solve the exam is one week.

4. STYLE INSTRUCTIONS: Read carefully before answering. Write the answers with order and without orthographic mistakes. Use all conceptual tools seen in class. The exam is individual.

¡SUCCESS IN YOUR EVALUATION!

Problem Setting:

You are working in a transnational UK e-commerce company as a Data Scientist and Machine Learning Specialist. Your job consists in analyzing the transactions made from 2010 and 2011 to gather business insights that will allow the company take better decisions.

All data was collected from the platform Kaggle in the following website https://www.kaggle.com/carrie1/ecommerce-data and was built by the UCI Machine Learning Repository at http://archive.ics.uci.edu/ml/index.php.

The data comprises invoices emitted by different customers, where each invoice may include one or more products with a unique identifier. The quantity and unit price are specified for each product included in the invoice. It is important to note that many invoices may have been cancelled and are identified with a "*C*" at the start of their corresponding "*InvoiceNo*". Furthermore, the invoice dates are specified using the US date format convention: *"mm/dd/yyyy"*. Moreover, the data includes the country where the invoice was emitted, and the description associated to each product.

Therefore, the company asks you to make the following analyses:

1. **Predict total sales per month for next year**

   Predict total sales for all months in the next year (2012). The company requires that you provide an estimate with 95.45% of confidence. The following figure shows an example of a prediction (this might not represent the real solution of the problem).

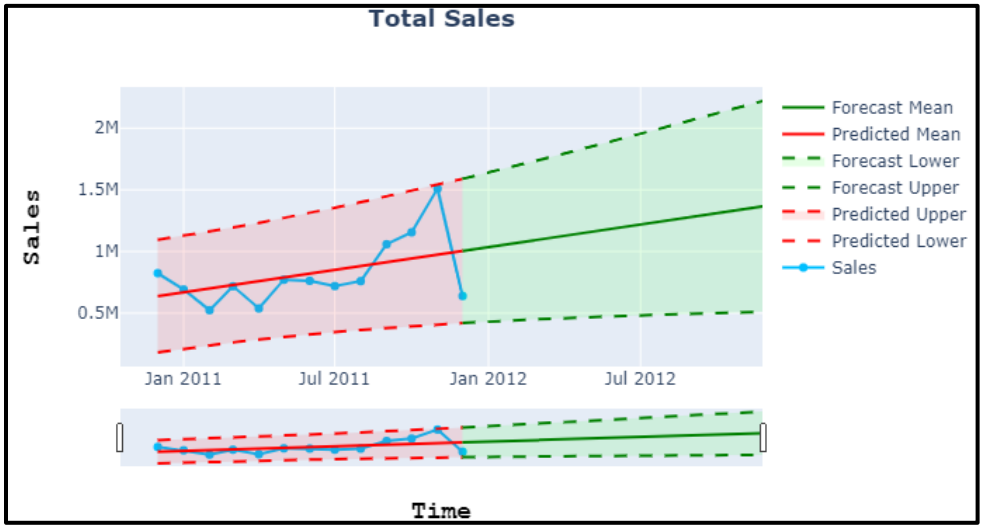

Figure 1. Example of a sales prediction curve

2. Predict the demand of the most requested product for next year

The company wants to know which is the most requested product. Furthermore, it wants to predict the demand of the product per month, which will allow it to anticipate customer needs for the next year. Again, your manager is requesting a prediction for the next year with a 95.45% confidence. The following figure shows an example of a prediction (this might not represent the real solution of the problem).

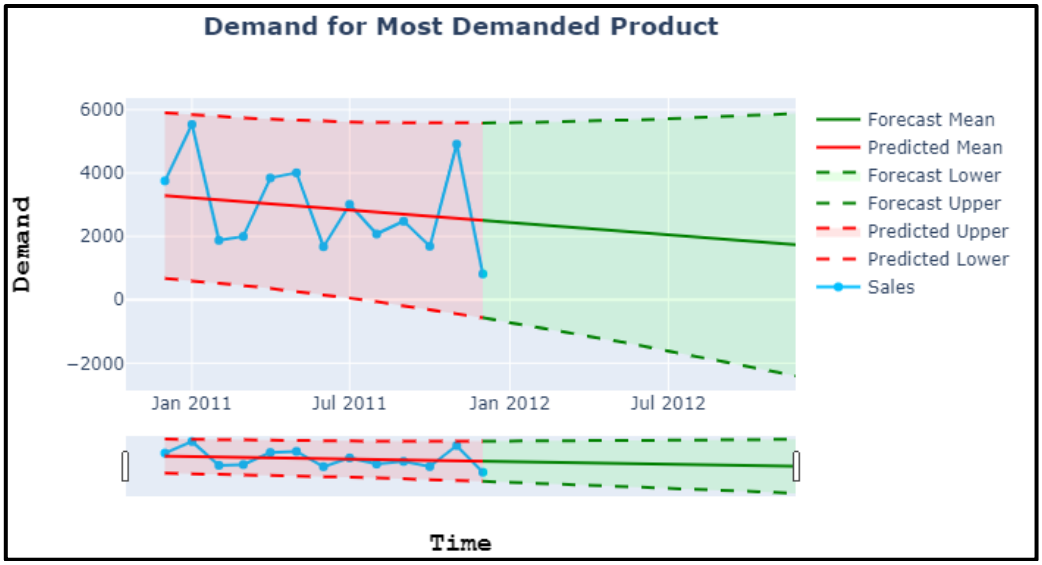

Figure 2. Example of a demand prediction curve.

3. Predict when the top 3 clients will call on a specific date interval

Finally, the company wants to know who the three top customers are based on the number of invoices emitted in the database. Furthermore, it requests you to make predictions to determine the probability that each one of the three top customers will call on a specific date interval. The following figure shows an example of the desired outcome (this might not represent the real solution of the problem).

```
Probability that customer 1 will call between 2011-12-13 and 2011-12-17 = 0.11
Probability that customer 2 will call between 2011-12-13 and 2011-12-17 = 0.10
Probability that customer 3 will call between 2011-12-07 and 2011-12-11 = 0.41
```

Figure 3. Example of call predictions for three customers.