
## Objective of Tech Track (Machine Learning)

# Build a Machine Learning model to predict the date of payment of an invoice.

## Business Overview

### Introduction to B2B Operations:

The B2B world operates differently from the B2C or C2C world. Businesses work with other businesses on **credit**. When a **buyer business** orders goods from the **seller business**, the **seller business issues an invoice for the same**. This invoice for the goods contains various information like the details of the goods purchased and when it should be paid.

The simplest definition of **accounts receivable** is money owed to an entity by its customers. Correspondingly, the amount not yet received is **credit** and, of course, the amount still owed past the due date is **Account receivables.** A more formal definition of A/R is:

> ***"Accounts Receivable represents money owed by entities to the firm on the sale of products or services on credit. In most business entities, accounts receivable is typically executed by generating an invoice and either mailing or electronically delivering it to the customer, who, in turn, must pay it within an established timeframe, called credit terms or payment terms."***

Seller business interacts with various businesses and sells goods to all of them at various times. Hence, the seller business needs to keep track of the total amount it owes from all the buyers. This involves keeping track of all invoices from all the buyers. Each invoice will have various important fields like a payment due date, invoice date, invoice amount, baseline date etc.

The buyer business needs to clear its amount due before the due date. However, in real-world scenarios , the invoices are not always cleared ie. paid in full amount by the due date. The date on which a customer clears the payment for an invoice is called the **payment date**.

### Account receivables Department:

1. In the ideal world, the buyer business should payback within the stipulated time (ie the **Payment Term**). However, in the real world, the buyer business seldom pays within their established time frame, and this is where the Account receivables Department comes into picture.
2. Every business consists of a dedicated Account receivables Department to collect and track payment of invoices.
3. It consists of a Account receivables team that is responsible for:

- Collecting payments from customers for their past due invoices
- Sending reminders and follow ups to the customers for payments to be made
- Looking after the entire process of getting the cash inflow
- Help the company get paid for the services and products supplied.

**For example:**
Let's suppose Walmart wishes to add 20,000 Nike shoes to its sports department. Now, Walmart becomes the buyer business whereas Nike is the seller business. Since this is B2B, the business is done on credit.

After successful delivery of the shoes, Nike establishes a Payment term within which Walmart must pay it's total open amount, let's suppose $1.2M.

If Walmart is unable to pay within the decided payment term, it becomes the duty of the Account receivables department of Nike to follow up and get Walmart to pay the outstanding amount.

## Problem Statement for ML Model

The objective of the first half of the winter internship project is:

- To build a Machine Learning Model to **predict the payment date of an invoice** when it gets created in the system.
- **Categorize** the invoice **into different buckets based on predicted payment date**.

You will be receiving an invoices dataset that contains the past payment information and behaviour of various buyers. Based on the previous payment patterns, the ML model needs to predict what will be the date a payment is made by the customer for an invoice.
The model also needs to predict which aging bucket the invoice falls into based on the predicted payment date.

**For example:**

| Invoice No | Due Date | Actual Open Amt($) | Payment Date | Aging Bucket |
|---|---|---|---|---|
| 1001 | 03/12/2020 | 9000 | 14/02/2021 | > 60 days |
| 1002 | 10/12/2020 | 5000 | 23/12/2020 | 0-15 days |
| 1003 | 17/12/2020 | 3000 | 16/01/2021 | 31-45 days |

In this example, the payment for the first invoice (1001) is due on 3rd December 2020. The payment for the invoice was actually made on 14th February 2021 so it was 73 days past due thus falling in the beyond 60 days aging bucket. Similarly for the second invoice(1002) which was due on 10th December 2020, the payment was actually made on 23rd December 2020, 13 days after the original due date thus it belonged in the 0-15 days aging bucket.
So the objective is to predict the payment date i.e. for the third invoice(1003), you have to build a model that predicts it's payment date as 16th January 2021 and thus categorizes it in the 31-45 days aging bucket.

## HIGH LEVEL REQUIREMENTS OF APPLICATION

Specifically, below are the major aspects of the application that needs to be developed. The details for each of the below is provided in the functional overview section.

1. **Invoices Dataset:**
    a. HighRadius will provide you a **invoices dataset** which you need to parse and process.
2. **AI Support in the application:**
    a. Add support for predicting the payment date for invoice(s).
    b. UI should have a button to trigger the prediction of payment date.
    c. Payment date needs to be persisted across sessions in the UI.

# FUNCTIONAL OVERVIEW

## (1) Invoices Dataset:

Below is the sample CSV file screenshot.

| account_id | document_numb | company_code | fiscal_year | branch | customer_numbe | fk_customer_ma | document_date_ | baseline_date_n | due_date_ |
|---|---|---|---|---|---|---|---|---|---|
| 60 | 20000950 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000950 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000950 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000950 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000870 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000870 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000870 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000870 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20001004 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20001004 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20001004 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20001004 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000831 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000831 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000831 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20000831 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |
| 60 | 20103019 | \N | \N | \N | 999888 | -1 | 11-04-16 | \N | \N |
| 60 | 20103019 | \N | \N | \N | 999888 | -1 | 11-04-16 | \N | \N |
| 60 | 20103019 | \N | \N | \N | 999888 | -1 | 11-04-16 | \N | \N |
| 60 | 20103019 | \N | \N | \N | 999888 | -1 | 11-04-16 | \N | \N |
| 60 | 20000972 | \N | \N | \N | 999888 | -1 | 05-04-16 | \N | \N |

**List of all the fields part of dataset are as follows:**

- Acct Doc ID
- Company ID
- Document Number
- Document Number Norm
- Business Code
- Create Year
- Document type
- Customer Number
- Customer Number Norm
- Customer Map ID
- Customer Name
- Document Create Date
- Document Create Date Norm
- Posting date

- Posting date norm
- Posting ID
- Due In Date
- Due In Date Norm
- Order Create Date
- Order Create Date Norm
- Invoice ID
- Invoice ID Norm
- Baseline Create Date
- Invoice Date norm
- Total Open Amount
- Customer Payment Terms
- Shipping date
- Shipping to
- Clear Date
- Clear Date Norm
- Is Open Invoice
- Doc Id
- Actual Outstanding amount
- Invoice Amount
- Dispute Valid Status
- Dispute Amount

Refer to this link for detailed description of column headers.

## (2) AI Prediction of predicted date of payment:

1. As part of the problem statement, the Payment date of an invoice will be predicted.
2. The **Predict button** will be located towards the left and over the invoices grid.
3. Clicking on the Predict button will populate the **Predicted Payment Date and the Predicted Aging Bucket** column of the grid with the predicted values derived from the ML model. The different buckets will be :
    - 0-15 days
    - 16-30 days
    - 31-45 days
    - 46-60 days
    - Greater than 60 days

## Glossary

1. Invoice - A document which is issued by a seller to a buyer when some goods are purchased. The fields which can be part of the invoice are defined below
2. Open Invoice - Invoices which are not cleared (payment not done) are called Open Invoices.
3. Closed Invoice - Invoices which are cleared (payment done) are called Closed Invoices.
4. B2B - Business to Business
5. B2C- Business to Consumer
6. C2C - Consumer to Consumer
7. Payment Terms - These indicate the period within which payments should be made and how. These terms are usually included in the invoices generated by companies and sent to customers.Eg Net 30, Net 60