4/18/2023

Bello,Mary

NSCC

**Predictive analysis for the wholesale market price for Corn.**

Project Document

Contents

[Executive Summary 2](#_Toc132754346)

[Final Business Case Documentation: 2](#_Toc132754347)

[Data source 2](#_Toc132754348)

[Overview of the project: 2](#_Toc132754349)

[Screenshot of ERD 3](#_Toc132754350)

[DFD (Level 0 and Level 1) 4](#_Toc132754351)

[Data Dictionary 5](#_Toc132754352)

[Code used to create your Database (DDL) 5](#_Toc132754353)

[Gantt Chart 6](#_Toc132754354)

[Screenshots of MS Power BI and ML Code 6](#_Toc132754355)

[Screenshots (png) of the following MS Power BI pages: Dashboard showing the KPIs you identified in your Business Case document. 6](#_Toc132754356)

[Descriptive Statistics related to your data. 7](#_Toc132754357)

[A prediction using the latest data. 7](#_Toc132754358)

[The latest Prediction Model Results showing RSME. 8](#_Toc132754359)

[A Report that will allow the user to filter (Slice/Dice) multidimensional data 8](#_Toc132754360)

[MS SSIS Packages/Windows Schedule (png) for automating the following: 9](#_Toc132754361)

[Screenshot of data import procedure (SSIS Package) 9](#_Toc132754362)

[Screenshot showing data download in the database. 9](#_Toc132754363)

[Automated ML Model builder (SSIS Package) 10](#_Toc132754364)

[Screenshot of Model running in SSIS 10](#_Toc132754365)

[Automation of the Model using Window Scheduler 10](#_Toc132754366)

[Screenshot showing automation complete. 11](#_Toc132754367)

[PowerShell script used to download the data from the Web. 11](#_Toc132754368)

[R Scrip used to build and save your prediction model. 11](#_Toc132754369)

[Screenshots (png) of the Data Import and ML Model builder Automation 13](#_Toc132754370)

[Screenshot showing Data Import automation 13](#_Toc132754371)

[Screenshot showing ML automation. 13](#_Toc132754372)

[Screenshot showing task complete. 14](#_Toc132754373)

# Executive Summary

AGRO farms Canada Inc deals on agricultural produce, the wholesale market price of agricultural produce is a critical factor in the success of the agricultural industry. This price is influenced by a variety of factors such as weather patterns, global demand, and supply chain disruptions. Knowing the future prices of agricultural produce can provide companies with a competitive advantage in the market.

This business case proposes the implementation of a predictive analytics solution for forecasting the wholesale market price of agricultural produce. The solution will use machine learning algorithms to analyze historical data and other relevant factors to make accurate predictions about future prices. The implementation of the solution will provide our company with valuable insights into the future prices of produce, allowing us to make informed decisions about purchasing, sales, and marketing strategies.

The implementation of this solution will result in improved competitiveness in the market, increased profitability and reduced risk, and better ability to respond to changes in market conditions. The solution will involve data collection, data preparation, model training, and deployment.

In conclusion, this predictive analytics solution will provide a competitive advantage for our company in the agricultural industry and will contribute to increased profitability and improved decision-making.

# Final Business Case Documentation:

This business case proposes the implementation of a predictive analytics solution for forecasting the wholesale market price of corn. The solution will use machine learning algorithms to analyze historical data to make accurate predictions about future prices. The implementation of the solution will provide our company with valuable insights into the future prices of produce, allowing us to make informed decisions about purchasing, sales, and marketing strategies.

# Data source

My data source is from Government of Canada open data, <https://open.canada.ca/data/en/dataset/b52424ca-ebb8-4847-808e-d525120b55df>

Each role of dataset reports the weekly wholesale market prices (highest and lowest) of a package and grade of specific variety of a commodity and from originating country/province/state. The Publisher is Agriculture and Agri-food Canada, and the Section name is Market Analysis and Information.

# Overview of the project:

I developed a business case, I created a Data Flow Diagram (DFD), Entity Relationship Diagram (ERD), Data Dictionary, and wrote a script that created a Data Definition Language (DDL), that creates a table in MS SQL server. For this project I used a Fact table.

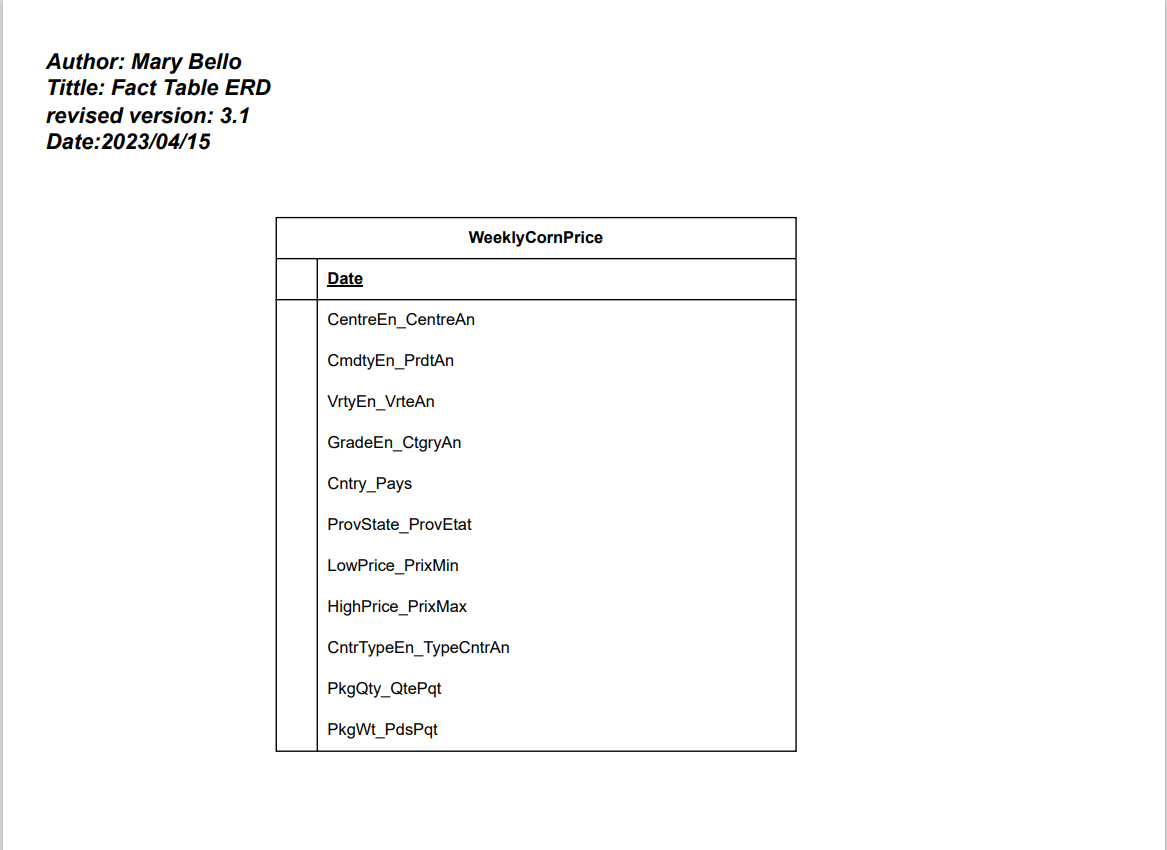
I used PowerShell to download the data from the government open data source and used SSIS to import the data into MS SQL server. In SQL, I cleaned my data, created a view t which I used for the actual modelling.

I then created wrote a script in R that creates a Model that will train the data that will be used for the machine learning (ML) and saved the script on my C drive as a variable which I will use to do the actual prediction. I automated data import process by deploying the SSIS package to the SQL.

I automated the data in MS SQL by creating my credentials, proxy, schedule and ran the job and I automated the Model by running it in a window’s scheduler.

Finally, I imported my data into Power BI to show the descriptive statistics of my data, display the RMSE, MAE and R2 of my model, and some visuals that will allow the user to filter multidimensional data.

## Screenshot of ERD



## DFD (Level 0 and Level 1)

Diagram

Description automatically generated

## Data Dictionary

Table

Description automatically generated

## Code used to create your Database (DDL)

Graphical user interface, text, application

Description automatically generated

## Gantt Chart

Timeline

Description automatically generated

# Screenshots of MS Power BI and ML Code

## Screenshots (png) of the following MS Power BI pages: Dashboard showing the KPIs you identified in your Business Case document.

Diagram

Description automatically generated

## Descriptive Statistics related to your data.

Chart, histogram

Description automatically generated

## A prediction using the latest data.

Table

Description automatically generated with medium confidence

## The latest Prediction Model Results showing RSME.

Graphical user interface

Description automatically generated with medium confidence

## A Report that will allow the user to filter (Slice/Dice) multidimensional data

Graphical user interface

Description automatically generated

# MS SSIS Packages/Windows Schedule (png) for automating the following:

## Screenshot of data import procedure (SSIS Package)

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated

## Screenshot showing data download in the database.

A screenshot of a computer

Description automatically generated with medium confidence

# Automated ML Model builder (SSIS Package)

## Screenshot of Model running in SSIS

Graphical user interface

Description automatically generated with low confidence

# Automation of the Model using Window Scheduler

Graphical user interface, text, application, email

Description automatically generated

## Screenshot showing automation complete.

Graphical user interface, text, application, email

Description automatically generated

## PowerShell script used to download the data from the Web.

Graphical user interface, text, application, email

Description automatically generated

## R Scrip used to build and save your prediction model.

#Install packages and load libraries

# Package names

packages <- c("ggplot2", "RODBC", "car", "caret", "corrgram", "gridExtra", "forecast")

# Install packages not yet installed

installed\_packages <- packages %in% rownames(installed.packages())

if (any(installed\_packages == FALSE)) {

install.packages(packages[!installed\_packages])

}

# Packages loading

invisible(lapply(packages, library, character.only = TRUE))

#Connecting to SQL, server = serverName and database=databaseName.

dbconnection <-odbcDriverConnect('driver={SQL Server};server=MSI\\SQLSERVER;database=PredictionDB;trusted\_connection=true')

#odbcClose(dbconnection)

#load data from Database using sql query

CornPrice <- sqlQuery(dbconnection,paste("SELECT \* FROM ModellingData;"))

# Close the database connection

odbcClose(dbconnection)

#Check levels

# Check number of unique levels in categorical variables

cat\_vars <- c("CentreEn\_CentreAn", "CmdtyEn\_PrdtAn", "VrtyEn\_VrteAn", "GradeEn\_CtgryAn", "Cntry\_Pays")

for (var in cat\_vars) {

print(paste(var, nlevels(as.factor(CornPrice[[var]]))))

}

CornPrice <- subset(CornPrice, select = -c(CmdtyEn\_PrdtAn))

CornPrice<- na.omit(CornPrice)

# Remove Date variable

CornPrice$Date <- NULL

# Model data using 80/20 split

index <- createDataPartition(CornPrice$Lead\_LowPrice\_PrixMin, p=0.8, list=FALSE)

# Subset training set with index

CornPrice.training <- CornPrice[index,]

# Subset test set with index

CornPrice.test <- CornPrice[-index,]

## 10-fold CV

# possible values: boot", "boot632", "cv", "repeatedcv", "LOOCV", "LGOCV"

CornPrice$Date <- NULL # Remove Date variable

CornPrice<- na.omit(CornPrice) #Remove all NA from data set

fitControl <- trainControl(method = "repeatedcv", number = 10, repeats = 10)

modelLR.cv <- train(Lead\_LowPrice\_PrixMin ~ ., data = CornPrice.training, method = "ridge", trControl = fitControl)

modelLR.cv

saveRDS(modelLR.cv, "C://Users//maryb//Desktop//AppliedProject//Corn//modelLR.cv.rds")

# Screenshots (png) of the Data Import and ML Model builder Automation

## Screenshot showing Data Import automation

Graphical user interface, application

Description automatically generated

## Screenshot showing ML automation.

Graphical user interface, text, application, email

Description automatically generated

## Screenshot showing task complete.

Graphical user interface, text, application, email

Description automatically generated