Dairy Cows Data Transformation Software: From Report Analysis to Performance Visualization

Juan David Pérez
Master of Engineering in Computer Engineering (M.Eng.Cp.E.)
Alfredo Cruz, PhD
Electrical and Computer Engineering and Computer Science Department
Polytechnic University of Puerto Rico

Abstract *– There is a big opportunity to apply technology and data analytics to the agriculture sector, specifically, to dairy farms. Dairy farmers are looking for solutions that helps them to increase their profitability, efficiency, and reduce labor variance. The Dairy Cows Data Transformation Software helps dairy farmers in the analysis of the data being provided by DHIA report. The application shows all the KPI to the farmer in a web-based dashboard fashion. This eliminates the need of hour of data analysis and manual calculations. The pilot implementation was performed with the support of the dairy farm "El Remanso" in Camuy, Puerto Rico.*

Key terms *– Data Analytics, Industry 4.0, Python, and Reporting*

INTRODUCTION

Nowadays technology is everywhere. You can find it in your car, house, phones, work, etc. Maybe, the question is: where technology is not creating impact as it should? The revolution of technology in manufacturing started during the end of 20th century and continued its evolution during the 21st century. It was called the Industry 3.0 ERA. Computers and automation took control of the manufacturing of cars, drugs, and even agriculture in a reduced magnitude. Industry 3.0 brings with it a ton of data from field devices to automated systems. A decade ago, the Industry 4.0 revolution was born. This era is dealing with use of the data generated by the technology and automation to use it for analytics, predictive decisions, correlation, and so on. In summary, the researcher wants to use the data available to improve farmers' processes and be more efficient with less. Industrial Manufacturing is already taking benefits from the use of data.

Dairy farms are the biggest single part of Puerto Rico's agricultural economy, accountable for 37% of total agricultural production or $390 million, in part, because the Island has set limits on fresh milk imports. There are around 266 dairy farms around the island averaging 200-240 cows per farm [1]. Figure 1 shows dairy farms distribution around the island.

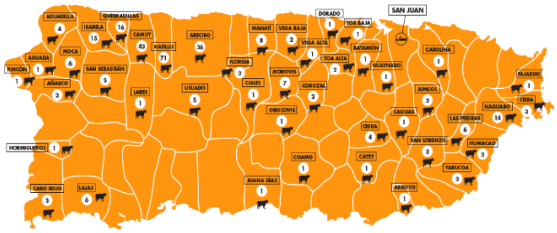

Figure 1

Dairy Farms in Puerto Rico

In total, dairy farms produce an estimate of 270 million quarts of milk per year. Most of the farms uses herringbone parlors with limited visibility on the performance of each individual cow due to lack of milk quality and yield sensors. Farmers relies on a Dairy Herd Improvement Association (DHIA) technician for a monthly average count of each cow's milk somatic cell (SCC) and milking yield among other indications [2]. The report generated as part of the technician visit is send electronically to the farmer. These reports are a valuable tool to monitor herd performance, sickness, conception rates, among many other KPIs.

This paper will summarize the current process used by dairy farmers to analyze, use, and manage the DHI reports and will go into details of the solution developed to minimize their effort and get

more value from the reports they received and analyzed.

PROBLEM STATEMENT

Dairy farmers face daily challenges to run their business. They need to constantly monitor the health of the cows and calves, milk quality, nutrition (feed and water), cow's welfare, reproduction rates, and, finally, the economics and profitability of the business. Technology is moving its way into dairy farms but at a slow rate, mainly due to the high cost of technology.

Nowadays, dairy farmers rely on monthly reports to gather important information about their herd. These reports are managed by the DHIA. DHIA sends a technician every 30 days to gather data of conception, cull, milk yield, and milk quality. Once all this data is gathered, DHIA will insert this data into their online database system and create a report that is send electronically to the farmer. This report consists of a Herd Summary and a Monthly summary section. In the Herd Summary section, there is information about feed cost, reproductive summary, birth summary, reproductive summary, and SCC, among many other KPIs. Figure 2 shows a portion of the monthly report and its data.

Raza
XX
Tipo Record
2-DHI-APCS

INFORME MENSUAL

Prueba Anterior	Fecha Prueba	Fecha Enviado
08-28-2019	09-25-2019	09-26-2019

Datos Día de Prueba

Raza	ID del Padre / ID Permanente	Leche Diaria: Prueba Anterior	Leche Diaria: Prueba Actual	% Pro / % Grasa	IOFC Diaria: Concentrados Actual	IOFC Diaria: Concentrados Indicado	SCC: Prueba Anterior	SCC: Prueba Actual	Estado: Código / Fecha	Índice Vaca: Ident. Vaca	Días en Prod.	Leche
HO	94WBN8173	56.0	47.0	3.2 / 2.7			107	81	2 / 03-24	107	186	12792
HO	94WBN8172	50.0	33.0	3.1 / 2.7			1056	1056	2 / 03-04	108	206	10198
HO	94WBO7262	66.0	58.0	3.1 / 2.6			35	187	2 / 07-29	112	59	3554
HO	94WBQ1700	16.0	18.0	3.1 / 2.6			17	71	2 / 02-10	200	228	9744
HO	94WBQ1624	16.0	13.0	3.1 / 2.6			528	62	2 / 02-11	201	227	7277

Figure 2

Monthly Herd Report

As previously mentioned, farmer use this report to make critical decisions related to production, Herd health, conception, and udder health. In a report written by Croushore [3], it states that DHIA testing is often underutilized on most dairy farms. Turning the raw data into something useful is important to take full advantage of the value of testing. Knowing what to do with the data; however, can be a little tricky. Udder health and mastitis, production and reproduction data are important information that will give the farmer the insight on how its business is running but more important, how is trending to detect anomalies that can impact the herd in the future. Managing all this data, coming from the report manually, is minimizing the benefits of this data to the farmer.

Project's Goals

Project goals are to develop a web-based dashboard with data extracted from the DHIA Monthly report. Some of the KPIs that can be collected from the pdf monthly report are:

- Milk per Cow per day
- Cows with high yield
- Cows with less than 20 liters (cows to dry)
- Days milked, Days Dry, Dates Open
- List of Cows to breed
- List of cows to dry
- Reproduction Data
- Breed or Heat Date
- Due Date
- Summary per Breed (Holstein, Jersey, Swiss, etc.)
- Health Data
- SCC Count per Cow analysis
- Higher cows
- Month to month comparison

Figure 3 illustrates how this project uses the DHIA data systematically.

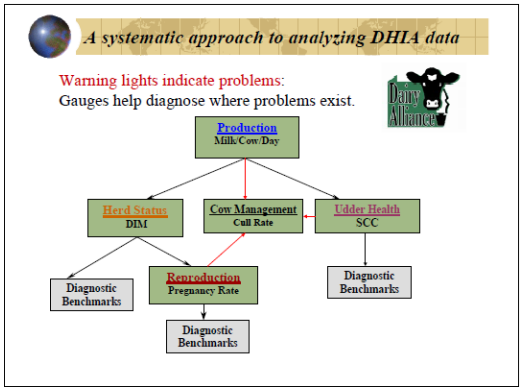

Figure 3
DHIA Systematic Approach [4]

With the use of python as a backend, the application transforms the DHIA pdf reports to a HTML file to "digitalize" the data in the report. Once the report is converted to HTML, the application lookup the data needed to generate the KPIs. This data is saved to a MySQL instance running in Azure to have the flexibility to run in the cloud for future enhancements.

Research Questions

The following questions were used as baseline and guideline for the development of the application:

1. How will the application help dairy cows and their farmers?
2. Is the data relevant to multiple users (dairy farmers)?
3. Will the farmer trust the system analytics to benefit from the manual data analysis?
4. Are dairy farmers in general willing to pay for this data analytics application?
5. How can this application be used to predict cow's health issues and improve pregnancy rates?
6. How many months the system will hold on to record?

Relevance and Significance

The main value of Dairy Herd Improvement records (DHIA, Tester-Sampler, and Owner-Sampler) comes from putting them to work. DHIA records provide abundant data for analyzing herd performance. However, many producers experience information overload when they receive their monthly reports. The result is information that is seldom or never used for its intended purposes: management of the dairy herd. Having a systematic approach to analyzing DHIA data can greatly improve dairy producer's understanding of dairy records and will increase their use of these records in making management decisions [4].

The KPIs allows producers to obtain a broad overview of herd performance in several key management areas: production, lactation and reproduction status, reproduction performance, udder health and overall cow management. It can compare the KPIs to the warning light on the dashboard of a car. It can tell the user something may be wrong, but it tells little about what may be causing the problem. To determine where problems may exist in the production system, the researcher need more detailed information. Can use diagnostic indicators to evaluate the various management areas of the overall production system.

KPIs can be retrieved from the monthly DHIA reports but not without hours of analysis. This ends up creating human errors that may affect the quality of the milk or the health of the herd. Not every farmer is using DHIA report as it should be. Most farmers only look for the herd summary page to understand how the herd is compared to previous months. They are missing the opportunity to increase their possibilities of influence into the current month performance by looking into KPIs that trigger immediate action thanks to the data analysis.

DHIA testing and its reports are not free of charge. Farmers paid up to $800 dollars monthly for the testing, laboratory analysis, and the resulting report. The Dairy Cows Data Transformation application adds more value to the service provided by DHIA. In 2018 a total of 15,558 farms were using DHIA (Refer to Figure 4). That is, 4.3 million cows recorded through the association. This demonstrates the value of the Dairy Cows Data Transformation application on helping these farmers to improve their herd performance.

Averages of DHI cow herds by herd size

	Records		Milk	Fat			Protein		
Herd size (cow-years per herd)	Herds	Cow-years	(lbs)	(%)	(lbs)	% Reporting	(%)	(lbs)	% Reporting
5- 24	314	5,040	16,804	4.11	689	100	3.24	543	100
25- 49	2,546	102,599	19,040	3.96	754	100	3.12	595	100
50- 74	3,459	209,657	20,953	3.90	817	100	3.11	652	100
75- 99	2,022	173,775	21,388	3.91	836	100	3.14	672	100
100- 149	2,208	268,604	21,913	3.91	857	99	3.14	689	99
150- 199	1,042	179,686	22,357	3.91	875	99	3.14	703	99
200- 299	1,141	277,878	23,369	3.88	909	99	3.14	736	99
300- 399	584	202,255	23,829	3.89	928	98	3.14	750	98
400- 499	395	176,465	25,049	3.87	969	98	3.13	783	97
500- 749	607	369,485	25,134	3.85	970	98	3.13	790	97
750- 999	318	275,484	25,578	3.85	987	98	3.14	805	97
1,000-1,499	353	431,397	25,198	3.87	979	98	3.15	798	96
1,500-1,999	199	342,665	25,903	3.85	997	100	3.16	820	99
2,000-2,999	195	475,165	25,839	3.88	1,004	100	3.18	823	98
3,000-3,999	83	283,311	25,258	3.88	982	98	3.20	808	97
4,000+	92	528,788	25,244	3.87	975	98	3.20	809	99
All Herds	15,558	4,302,253	24,324	3.88	944	99	3.16	769	98

Figure 4
Average DHI herds 2018 [2]

As part of the initial investigation of feasibility of this project, the project team got into contact with Jeffrey Bewley, PhD who is a Dairy Housing and Analytics Specialist for Alltech. He is a respected doctor with multiple journals related to dairy farming and cow's health. His impression of the Dairy Cows Data Transformation project was that it might have a potential to be on the market once this pilot is proved to be reliable and accurate in the data acquisition.

METHODOLOGY AND DESIGN

To help the farmer to maximize the value of the monthly report from DHIA, the Dairy Cows Data Transformation project was developed to automatically create dashboards and reports with data from each of the Key performance indicators (KPIs).

With this data extraction, the project improved the visibility of the farmer by extracting the most important information from the DHIA reports electronically and analyze it to show alerts to take the necessary actions. In addition, having the data digitalize and stored in a database provides the farmer flexibility to add additional metrics, reports, and dashboards in the future.

The application is structure using a database, a backend application for the extraction process, and a Web interface to upload the files to be analyze and display the dashboard and reports. In the next few paragraphs, I will go over each of these components of the project.

Database

For the data repository, the application is connecting to an Azure Workspace MySQL database. The software used to create and modified the database is MySQL Workbench. Figure 5 shows a snapshot of the connection used to open the DB to the application.

```
cnx = mysql.connector.connect(user='ideadmin@dhia', password='Piratas2022!',
                              host="dhia.mysql.database.azure.com",
                              port=3306,
                              database='dhia')
cursor = cnx.cursor()
```

Figure 5

Database Connection

The tables created in the database are used to store data from DHI report in pdf format and, later, extract the data for analytics. Figure 6 shows the tables created. The main table used to store monthly cow data is the dhia.cow_mon-thly_record. This table will be used by the Front-End reporting software to generate most of the visualization.

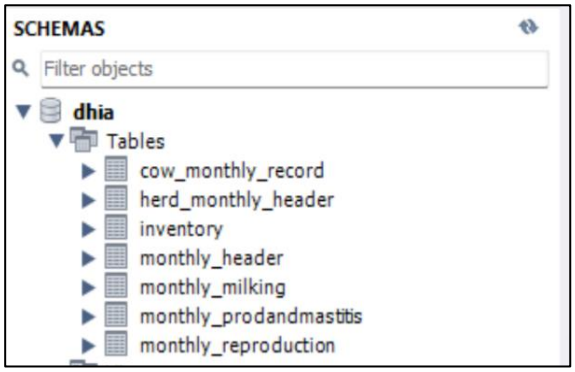

Figure 6
Database tables

An entity relationship diagram (ERD) was designed to better understand the relationship between each of the tables. Table *Monthly_hea-der* is linked to tables *cow_monthly_record* and herd_monthly_header using the table *month-ly_header* PK (primary key) "TestDate" as their FK (foreign key). Tables *monthly_pro-dandmastitis*, *monthly_reproduction*, and *month-ly_milking* are related to table *herd_month-ly_header* with the PK idherd_month-ly_header as their FK. Figure 7 shows the Entity Relationship Diagram for the Dairy Herd Management System where the graphical relationship between the tables (entities) can be observed.

Backend Software

The project team decided to use Python as the backend of the system. With python the team found flexibility in the packages available for pdf files conversion, database connectivity, and frontend with web functionality. Figure 8 shows portions of the code used to convert the pdf data into html code to be transferred later to the database.

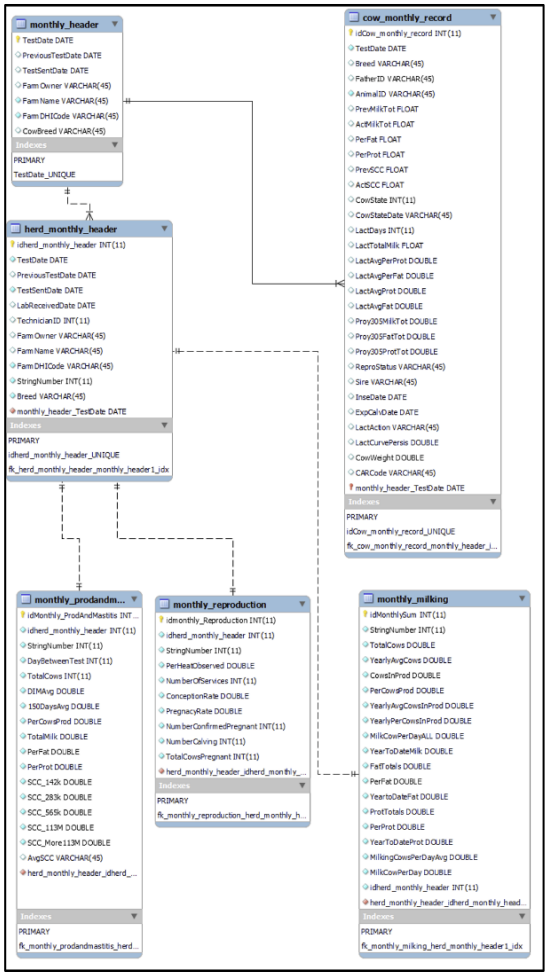

Figure 7

Dairy Herd Database ERD

```
output = StringIO()
with open(file_name, 'rb') as pdf_file:
    extract_text_to_fp(pdf_file, output, laparams=LAParams(),
                       output_type='html', codec=None)
with open('example8.html', 'w+') as html_file:
    html_file.truncate(0)
    html_file.seek(0)
    html_file.write(output.getvalue())
```

Figure 8

pdf to html conversion scripting

The process to gather the file starts at the Front-End Html page, where the user will choose the monthly pdf file and submit it to the application. The python code is continuously listening to the folder where the files are dropped. Once there is a new file, the program converts it to html to transform each different table in the report and encapsulate the data in specifics frames and tags that will help the system to know the exact location of each data value the team want to store.

Once the pdf file tables have been converted to an html file, the system will go through the html and retrieve the information needed to be stored in the database. This information will be used to create reports of the raw data stored. Figures 9 shows the acquisition and data storage of the header data used to fill the "monthly_header table.

```
a = True
while a:
    # read next line
    line = file1.readline()
    # check if line is not null
    if not line:
        a = False
    else:
        # you can access the line

        ############## Start of Father and Cow ID###############
        if flag5 and 'left:34px' not in line:
            flag5 = 0
        if flag6:  ##New Added
            flag6 = 0
            if 'px">' in line:
                TestPrint = line.partition('7px">')[2]
            else:
                TestPrint = line.partition('<br>')[2]
            IDList2.append(TestPrint.rstrip('\n'))
```

Figure 9

Data Retrieval from html to String Tags

Once all the data needed from the html file is extracted to the system, the application will execute an insert command to store the data in MySQL tables previously created. Figure 10 shows the insert instruction.

```
insert_Statement = """INSERT INTO `dhia`.`monthly_header`
(`TestDate`, `PreviousTestDate`, `TestSentDate`, `FarmOwner`,
`FarmName`, `FarmDHICode`, `CowBreed`)
VALUES
(%s,
%s,
%s,
%s,
%s,
%s,
%s);"""
IDList3 = (TestDate, PrevTest, SentDate, FarmerName,
           FarmName, FarmDHICode, Breed)
try:
    cursor.execute(insert_Statement, IDList3)
```

Figure 10

Insert Python Command

Upon completion of the insertion query without errors, the database tables will have all the

information needed for the Dashboard. Figure 11 shows the data inserted to the database.

TestDate	PreviousTestDate	TestSentDate	FarmOwner	FarmName
2019-07-31	2019-06-20	2019-08-01	RAFY LOPEZ	VAQ REMANSO
2019-08-28	2019-07-31	2019-09-04	RAFY LOPEZ	VAQ REMANSO
2019-09-25	2019-08-28	2019-09-26	RAFY LOPEZ	VAQ REMANSO
NULL	NULL	NULL	NULL	NULL

Figure 11

Database with data retrieved from pdf file

Web Interface

Web Interface was developed using the library Flask for Python. Flask is a Python framework for building web apps for machine learning and data science. Figure 12 shows the python instructions that opens the webserver port 5000, sets the maximum size of the file to be uploaded, limits the types of files to upload and, once the file is submitted by the user, it will send the file to its folder.

```
# initialising the flask app
app = Flask(__name__)

# Creating the upload folder
upload_folder = "uploads/"
if not os.path.exists(upload_folder):
    os.mkdir(upload_folder)

# Max size of the file
app.config['MAX_CONTENT_LENGTH'] = 1 * 1024 * 1024

# Configuring the upload folder
app.config['UPLOAD_FOLDER'] = upload_folder

# configuring the allowed extensions
allowed_extensions = ['pdf']
```

Figure 12

Webserver & Initial Page Setup

The script shown in Figure 13 will create portions of the webpage dynamically. This functionality helps reduce coding and allows create web pages with table and analytics.

Figure 14 shows an abstract of the webpage dashboard showing the upload file button, the submit button and portion of the production per cow comparison between previous month and current month. This dashboard is just a simple report providing value to the farmer.

```
@app.route('/upload', methods=['GET', 'POST'])
def uploadfile():
    if request.method == 'POST':  # check if the method is post
        f = request.files['file']  # get the file from the file
        # Saving the file in the required destination
        if check_file_extension(f.filename):
            f.save(os.path.join(app.config['UPLOAD_FOLDER'],
                                secure_filename(f.filename)))
            return 'file uploaded successfully'  # Display this
        else:
            return 'The file extension is not allowed'
```

Figure 13

Webpage Dynamic Creation Script

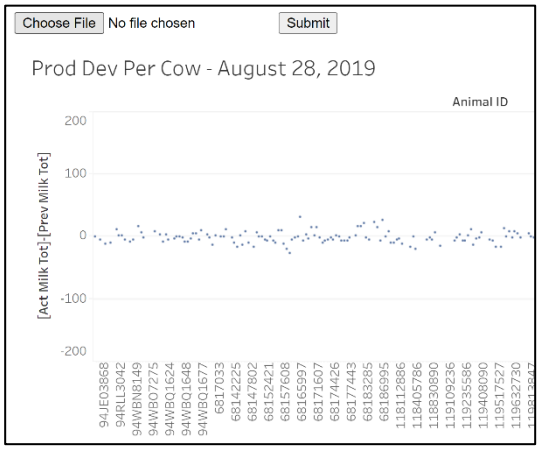

Figure 14

Web Application Main Page

Reporting and Visualization Tool

There are numerous reporting tools available to transform database data into meaningful visualization and dashboards. The one selected for this project is Tableau. This online/cloud tool connects to Azure MySQL through a JDBC driver and is flexible to allow the creation of multiple reports and visualization tables. Figure 15 shows the development screen of the application.

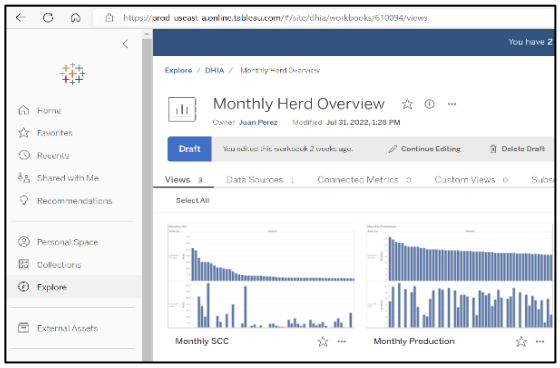

Figure 15

Tableau Development Environment

FUTURE WORK

The Dairy Cow Transformation Software have the capability of expand its functionalities beyond what it has now. This provides the opportunity of increasing the capabilities of the system by including additional KPIs and reports for the farmer. Also, the researchers can add the generation of automatic alerts to the farmer with the data entered to the system, not only from the current month but from previous months stored in the database. This will increase the system value to the clients with the help of data correlation. Moreover, with the data transformation from the monthly PDF files to MySQL database the researchers have the flexibility of adding more value with minimum effort in terms of software development.

CONCLUSIONS

The main value of Dairy Herd Improvement records (DHIA, Tester-Sampler, and Owner-Sampler) comes from putting them to work based on the information provided monthly. DHIA records provide an abundance of data for analyzing herd performance. However, many producers experience information overload when they receive their monthly reports. The result is information that is seldom or never used for its intended purposes which is the management of the dairy herd. Having a systematic approach with the use of the Dairy Cow Transformation Software to analyzing DHIA data will greatly improve dairy producer's understanding of dairy records and will increase their use of these records in making management decisions.

The KPIs allow producers to obtain a broader overview of herd performance in several key management areas: production, lactation and reproduction status, reproduction performance, udder health and overall cow management. The KPIs can be compared to the warning light on the dashboard of a car; it can tell us something may be wrong, but it tells a little about what may be causing the problem. To determine where problems may exist in the production system, farmers need more detailed information. That is where the value of the Dairy Cow Transformation Software relies: it creates performance indications to evaluate the various management areas of the overall production system and help farmers to make better decisions for their business.

ACKNOWLEDGEMENT

Thanks to my advisor Dr. Alfredo Cruz for his guidance, patience, and advise.

Thanks to Rafael López, owner of "El Remanso" farm and Dr. Guillermo Ortiz, Professor at UPRM Mayaguez for their guidance to understand dairy farmers KPIs and their challenges.

REFERENCES

[1] "Nuestros Ganaderos – Indulac," *Indulac*, 2022 [Online]. Available: https://indulacpr.com/nuestros-ganaderos/?lang=enstros-ganaderos%2F%3Flang.DHIA.

[2] "USDA summary of herd averages," *CDCB*, 2018 [Online]. Available: https://queries.uscdcb.com/publish/dhi/current/haall.html

[3] B. Croushore, D. W. O. Clinic, and PC, "What to do with DHIA data," Daily American, *The Daily American*, 02-Feb-2017 [Online]. Available: https://www.dailyamerican.com/story/lifestyle/home-garden/2017/02/02/what-to-do-with-dhia-data/44071501/

[4] Hilty, "Interpreting Your DHIA Records." *Penn State Dairy Alliance*, 2018.