

Rodrigo Rivas González

Mechatronic Engineer — Data & Business Analyst

Email: Rodrivas2003@gmail.com

LinkedIn: www.linkedin.com/in/rodrigo-rivas-gonzález

GitHub: [rodrivas2003.github.io](https://github.com/rodrivas2003)

About Me

Mechatronics Engineering student with a strong focus on data analytics, machine learning, control systems, and engineering design. I am currently pursuing a Bachelor's degree in Mechatronics Engineering, where I have developed a solid foundation in mechanical systems, electronics, control, and programming. I am particularly interested in applying data-driven approaches and control theory to real-world engineering problems, combining analytical modeling, simulation, and hands-on implementation to build high-performance systems.

Industrial Automation System for BIC Pen Testing Machine

Role: PLC Logic Programmer

Tools: Allen-Bradley PLC, Studio 5000, Ladder Logic, FactoryTalk View

What is it?

Mechatronic automation proposal made tailor-made for the company BIC, for an industrial testing machine designed to evaluate and classify BIC pens. The system performs sequential quality tests, classifies products as approved or rejected, and deposits accepted units automatically. The project integrates sensors, pneumatic actuators, PLC-based control logic, and an HMI for supervision.

My responsibility in the project

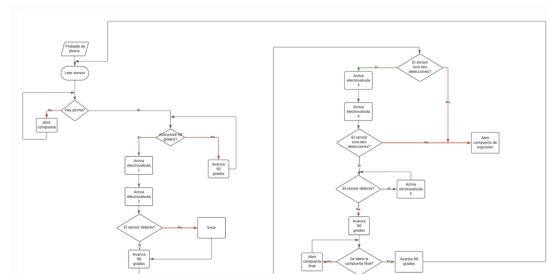
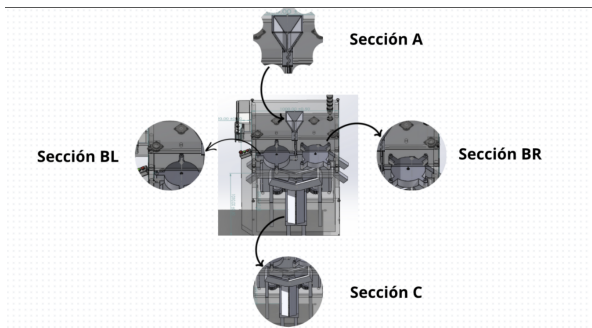
I was responsible for the development of the PLC logical control. This included designing the complete sequential logic of the machine, defining the execution order of each station, and ensuring safe and consistent operation through conditional logic, timers, flags, and resets. Additionally, I reviewed the mechanical design to verify logical-mechanical coherence between actuators, sensors, and control timing.

What I learned

Industrial PLC programming principles, sequential control strategies, and the importance of synchronizing mechanical movement with logical execution. I also gained experience in structuring scalable ladder programs and understanding real-world constraints caused by limited sensor feedback.

What I applied

Implementation of ladder logic using Studio 5000, including set/reset logic, timers, counters, and conditional task execution. I applied cascade-style sequencing triggered by a capacitive sensor as the system reference, ensuring correct transitions between testing stations and the deposit stage. Mechanical design review was used to validate timing assumptions and pneumatic actuation behavior.



Customer Segmentation and Demand Forecasting for NATA Supermarkets

Role: Business Analytics & Data Science Analyst **Tools:** Python, Pandas, Scikit-learn, Statsmodels, K-Means, Excel

What is it?

End-to-end business analytics case focused on improving customer segmentation, demand forecasting, and inventory planning for NATA Supermarkets. The project combines data preprocessing, unsupervised learning, time-series forecasting, and operational insights to support data-driven decision making.

My responsibility in the project

I was responsible for the data preparation pipeline, including cleaning, handling missing values, removing duplicates, and feature formatting. I also led the analytical decision for the number of clusters in the K-Means model by evaluating Elbow Method behavior and Silhouette Scores, and translating the technical results into a business-justified choice. Additionally, I developed the forecasting models and generated future demand predictions used as inputs for inventory analysis.

What I learned

How data quality directly impacts model performance, the trade-off between statistical optimality and business interpretability when selecting K in clustering, and the strengths of Double Exponential Smoothing for time series with trend components. I also learned how to evaluate forecast accuracy using RMSE and MAPE in a business context.

What I applied

Structured data-cleaning workflows, K-Means clustering with PCA visualization, analytical selection of K using Silhouette Score and Elbow Method, and demand forecasting using Double Exponential Smoothing. I applied error metrics (RMSE, MAPE) to compare models and produced multi-period predictions that directly informed downstream inventory and lot-sizing decisions.

DATA

Nata, in its attempt to improve, provided us with a database containing various factors they consider important; therefore, our work will depend on properly using this database in order to implement methodologies to address their issues.

The core problem being the failure when identifying purchasing patterns and customer preferences, improvement of promotion effectiveness and the forecasting the product demand more accurately.

STEP 1: DROP ROWS WITH ALL NANS

STEP 2: DROP COLUMNS WITH NAN PROPORTION > CLEAN_RATE

STEP 3: DROP DUPLICATE ROWS

STEP 4: FILL MISSING VALUES

STEP 5: CONVERT TO DATETIME

STEP 6: CHECK REMAINING MISSING VALUES

```
# Step 1: Drop rows where all values are NaN
df_clean = df.dropna(how='all')

# Step 2: Drop columns where NaN proportion is clean_rate
clean_rate = clean_rate * 1000 / 1000
df_clean = df_clean.drop(columns=[col for col in df_clean.columns if df_clean[col].isnull().sum() > (clean_rate)])

# Step 3: Drop duplicate rows
df_clean = df_clean.drop_duplicates(inplace=True)

# Step 4: Separate categorical and numerical columns
categorical_cols = df_clean.select_dtypes(include=['object']).columns
numerical_cols = df_clean.select_dtypes(include=['int64', 'float64']).columns

# Step 5: Fill missing values
df_clean[categorical_cols] = df_clean[categorical_cols].fillna('unknown')
df_clean[numerical_cols] = df_clean[numerical_cols].fillna(0)

# Step 6: Check remaining missing values
df_clean.isnull().sum()
```

HOW TO KNOW THE VALUE OF K

01 METHOD SELECTION

- Applied Elbow Method and Silhouette Score to determine optimal K
- Tested K values from 2 to 10

02 RESULTS ANALYSIS

- Elbow Method: Shows gradual decrease, no clear "elbow"
- Silhouette Score: K=2 achieved highest score (0.910)

03 FINAL DECISION

WHY K=3 INSTEAD OF K=2?
Better binary segmentation
Balanced distribution

K-means with K=3

Distribution of clusters:	
Cluster_K3	
0	1077
1	889
2	274
Name:	count, dtype: int64

TOO GENERIC

FORECASTING

DOUBLE EXPONENTIAL SMOOTHING

Moving Average fails to capture trends and reacts slowly to recent changes. Simple Exponential Smoothing only models the level but ignores trend components. Double Exponential Smoothing is superior because it captures both the current level and trend direction, providing more accurate multi-period forecasts for time series with fluctuating patterns like wine sales data.

```
from statsmodels.tsa.holtwinters import ExponentialSmoothing

forecast_periods = 4
results = {}

print("==== FORECAST WITH DOUBLE EXPONENTIAL SMOOTHING =====")

clusters = {
    0: df_cluster0,
    1: df_cluster1,
    2: df_cluster2
}

for cluster_name, df_cluster in clusters.items():
    print(f"--- Cluster '{cluster_name}' ---")
    cluster_results = {}

    for mkt_col in mkt_columns:
        print(f"--- Product: '{mkt_col}' ---")
        X = df_cluster[cluster_name][mkt_col].values
        y = df_cluster[cluster_name].values

    column_forecast = [cluster_0, df_customer, df_order, df_product, df_sales, df_time, df_location, df_promotion, df_discount, df_coupon, df_gift, df_reward, df_referral, df_affiliate, df_social, df_email, df_sms, df_push, df_inapp, df_app, df_browser, df_device, df_os, df_browser_version, df_device_model, df_device_manufacturer, df_device_brand, df_device_type, df_device_form_factor, df_device_screen_size, df_device_screen_resolution, df_device_screen_density, df_device_screen_orientation, df_device_screen_aspect_ratio, df_device_screen_refresh_rate, df_device_screen_color_depth, df_device_screen_brightness, df_device_screen_contrast, df_device_screen_gamma, df_device_screen_temperature, df_device_screen_humidity, df_device_screen_pressure, df_device_screen_vibration, df_device_screen_acceleration, df_device_screen_gyro, df_device_screen_magnetometer, df_device_screen_proximity, df_device_screen_compass, df_device_screen_barometer, df_device_screen_ambient_light_sensor, df_device_screen_fingerprint, df_device_screen_face_unlock, df_device_screen_voice_recognition, df_device_screen_biometric_security, df_device_screen_authentication_method, df_device_screen_authentication_strength, df_device_screen_authentication_frequency, df_device_screen_authentication_timeout, df_device_screen_authentication_attempts, df_device_screen_authentication_success_rate, df_device_screen_authentication_failure_rate, df_device_screen_authentication_error_rate, df_device_screen_authentication_warning_rate, df_device_screen_authentication_critical_rate, df_device_screen_authentication_fatal_rate, df_device_screen_authentication_abandonment_rate, df_device_screen_authentication_cancellation_rate, df_device_screen_authentication_revocation_rate, df_device_screen_authentication_deletion_rate, df_device_screen_authentication_expiration_rate, df_device_screen_authentication_expiry_date, df_device_screen_authentication_expiry_time, df_device_screen_authentication_expiry_timezone, df_device_screen_authentication_expiry_offset, df_device_screen_authentication_expiry_rule, df_device_screen_authentication_expiry_condition, df_device_screen_authentication_expiry_exception, df_device_screen_authentication_expiry_note, df_device_screen_authentication_expiry_comment, df_device_screen_authentication_expiry_metadata]
```

Observations = Months

What we took into account?

```
===== FORECAST WITH DOUBLE EXPONENTIAL SMOOTHING =====

CLUSTER 0 ---
Product: Metaxas
MSE: 10.000
RMSE: 3.162
Number of observations: 21
Forecast for next 4 months: [-1000.00, -1403.52, -1800.00, -1800.00]

Product: Metaxas
MSE: 10.000
RMSE: 3.162
Number of observations: 21
Forecast for next 4 months: [-1000.00, -1403.52, -1800.00, -1800.00]

Product: Metaxas
MSE: 10.000
RMSE: 3.162
Number of observations: 21
Forecast for next 4 months: [-1000.00, -1403.52, -1800.00, -1800.00]
```

sEMG Signal Amplifier and Conditioning Circuit

Role: Circuit Designer & Builder **Tools:** Operational Amplifiers (TL084, LM741), Analog Filters, Oscilloscope, Protoboard

What is it?

Design and implementation of an analog signal conditioning circuit to capture and process surface electromyography (sEMG) signals for biofeedback applications. The system amplifies, filters, rectifies, integrates, and conditions low-amplitude muscle signals to obtain a clean, interpretable output representing muscle activity.

My responsibility in the project

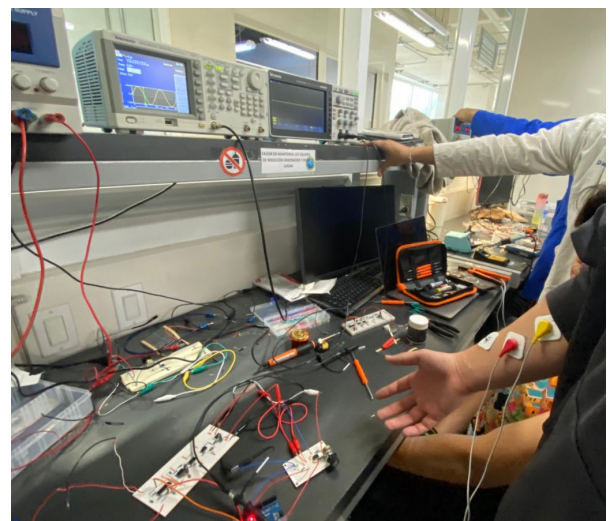
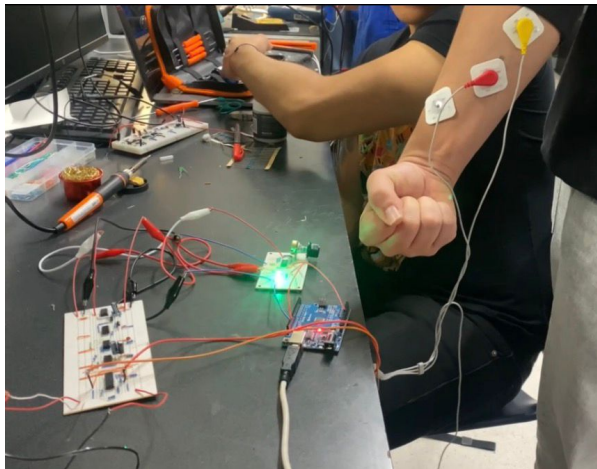
I was responsible for the complete circuit design and implementation. This included selecting all electronic components, designing each amplification and conditioning stage, performing theoretical calculations for gain, cutoff frequencies, and time constants, and physically building and wiring the circuit on a protoboard. I also validated each stage using an oscilloscope to ensure the measured behavior matched the expected analytical results.

What I learned

Analog signal processing for biomedical applications, including instrumentation amplifiers, band-pass filtering for noise reduction, rectification of AC biosignals, and signal integration. I gained strong intuition on how theoretical equations translate into real electrical behavior, and how component tolerances and noise affect low-amplitude biological signals.

What I applied

Design and calculation of an instrumentation amplifier using operational amplifiers with high CMRR, implementation of band-pass filters targeting the 20–500 Hz sEMG frequency range, precision rectification, and signal integration for envelope detection. I applied practical debugging techniques using an oscilloscope to verify gain, frequency response, and signal stability across all stages.



Maternal Health Risk Prediction Using Machine Learning

Role: Data Analyst & Machine Learning Contributor **Tools:** Python, Pandas, NumPy, Scikit-learn, Jupyter Notebook

What is it?

Machine learning project aimed at predicting maternal health risk levels (low, medium, high) during pregnancy using clinical and physiological indicators. The model supports early risk identification to assist healthcare professionals in preventive decision making.

My responsibility in the project

I was responsible for data cleaning and preprocessing, including handling missing values, removing duplicates, validating data types, and standardizing numerical features for model consistency. I also identified and treated outliers using statistical methods to prevent extreme values from biasing the model. Additionally, I collaborated with a teammate in the development and evaluation of the predictive model.

What I learned

The importance of data quality in healthcare-related machine learning projects, particularly how preprocessing decisions directly impact model reliability. I also gained experience in outlier detection techniques, feature scaling, and collaborative model development in a team-based data science environment.

What I applied

Structured data preprocessing workflows, Z-score-based outlier detection, feature standardization, and preparation of clean datasets for supervised learning. I contributed to model training and evaluation using classification metrics such as accuracy, precision, recall, and F1-score, supporting a Random Forest-based risk prediction approach.

Random Forest Results:				
	precision	recall	f1-score	support
0	0.92	0.82	0.86	120
1	0.80	0.82	0.81	101
2	0.82	0.93	0.87	73
accuracy			0.85	294
macro avg	0.84	0.86	0.85	294
weighted avg	0.85	0.85	0.85	294

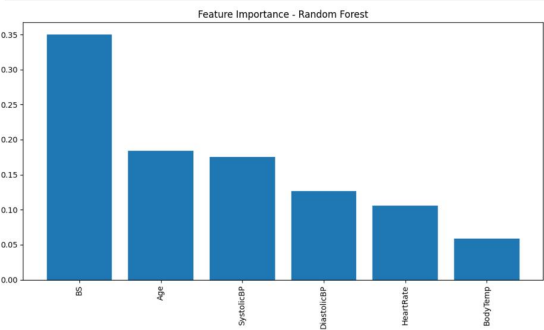
Accuracy: 0.8469387755102041

Confusion Matrix:

```
[[98 17  5]
 [ 8 83 10]
 [ 1  4 68]]
```

```
[33]: # Get feature importances
importances = rf.feature_importances_
indices = np.argsort(importances)[::-1]

# Plot top features
plt.figure(figsize=(10,6))
plt.title("Feature Importance - Random Forest")
plt.bar(range(X.shape[1]), importances[indices])
plt.xticks(range(X.shape[1]), X.columns[indices], rotation=90)
plt.tight_layout()
plt.show()
```



- Possible threshold: Age > 35.00 years → higher risk
- Possible threshold: SystolicBP > 120.00 mmHg → higher risk
- Possible threshold: DiastolicBP > 90.00 mmHg → higher risk
- Possible threshold: BS > 11.00 mmol/L → higher risk
- Possible threshold: BodyTemp > 98.00 °C → higher risk
- Possible threshold: HeartRate > 78.00 bpm → higher risk

Proposal for the Manufacturing of a Hydraulic Crane

Role: Mechanical Analysis & MATLAB Programmer
MATLAB, Engineering Mathematics

Tools: Free-Body Diagrams, Static Equilibrium,

What is it?

Engineering proposal for the manufacturing of a hydraulic workshop crane, focused on evaluating its structural behavior, internal forces, and piston loads under different lifting angles. The project combines analytical static analysis with computational simulation to validate the feasibility and safety of the proposed design.

My responsibility in the project

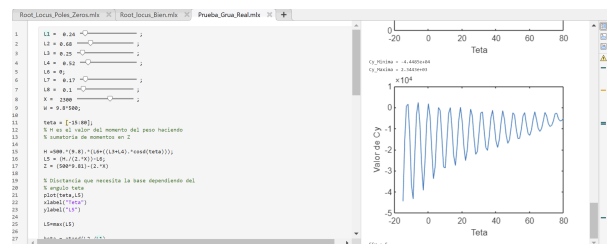
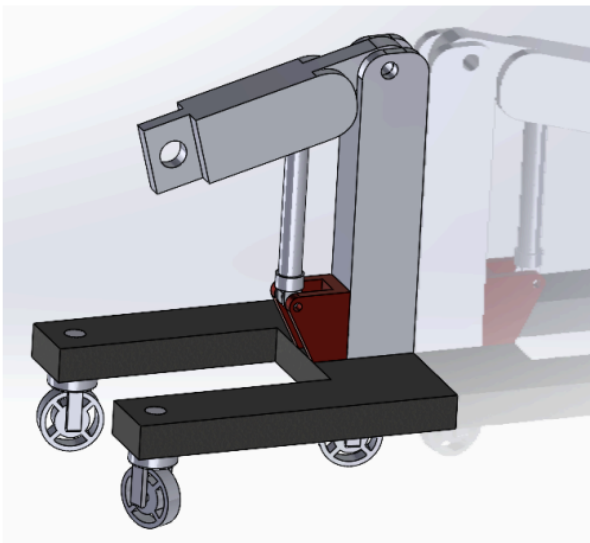
I was responsible for developing the free-body diagrams of the crane and its individual components, identifying two-force and three-force members. Based on these diagrams, I derived the static equilibrium equations and implemented a MATLAB program to solve the system numerically. The code evaluates piston force, reaction forces, and crane geometry across the full operating range.

What I learned

How to translate a mechanical concept into a manufacturable engineering proposal using static equilibrium analysis. I strengthened my ability to decompose complex mechanisms into free-body diagrams and convert analytical models into numerical simulations for design validation.

What I applied

Construction of detailed free-body diagrams, formulation of equilibrium equations in two and three dimensions, and numerical simulation in MATLAB. I performed parametric analysis over a range of lifting angles to determine piston length variation, reaction forces, and maximum load conditions, supporting the feasibility of the hydraulic crane design.



Design, Control, and Physical Implementation of a 2-DoF Robotic Arm

Role: Control Systems Engineer & Mechanical Assembly Lead **Tools:** Oscilloscope, MATLAB, Simulink, PID Control, Arduino, DC Motors with Encoders

What is it?

Development of a two-degrees-of-freedom (2-DoF) robotic arm combining experimental system identification, cascade PID control, trajectory generation, and physical construction. The system is capable of following predefined trajectories using position and speed control implemented through Simulink and embedded hardware.

My responsibility in the project

I was responsible for experimentally obtaining the transfer functions of the DC motors using an oscilloscope, extracting time constants and gains from step-response data. I contributed to the implementation of cascade PID controllers in Simulink and collaborated on tuning the control loops for speed and position. I also programmed trajectory generation blocks using MATLAB functions and inverse kinematics, and I was fully responsible for the physical construction and mechanical assembly of the robotic arm.

What I learned

How to bridge theoretical control design with real-world implementation by identifying system dynamics experimentally. I gained practical experience with cascade PID control, trajectory generation, and the impact of non-idealities such as friction, backlash, and sensor noise on robotic systems.

What I applied

Experimental system identification using oscilloscope measurements, first-order approximation of DC motor dynamics, implementation of cascade PID control in Simulink, and trajectory planning through inverse kinematics. I applied mechanical design and assembly principles to build a functional robotic arm capable of executing controlled movements and following geometric paths.

