Data Science Portfolio

By : Dinda Raraswati

SELF-OVERVIEW

Hi, I’m Dinda!

A data enthusiast with a Master’s degree in Agricultural Science, specializing in data science applications for agriculture.

Proficient in Python, SQL, and data visualization tools, with hands-on experience applying machine learning, time series analysis, and remote sensing for predictive insights.

Currently enhancing my skills through a Data Science Bootcamp at Dibimbing, I am equipped with both technical skills and business knowledge, and am now seeking to transition into a junior data scientist role where I can apply my analytical capabilities to solve impactful business problems.

EDUCATION

Bandung Institute of Technology (ITB)

Bachelor of Science in Agricultural Engineering
Aug 2016 – Jul 2020

- **Relevant Coursework :** Statistics for Agriculture, Introduction to Information Technology B
- **Thesis :** Quantification in The Change of Total Nitrogen in Soil and Plant Tissue of Black Soybean (Glycine soja L.) Detam 1 during The Cultivation in The Field
- **Experiences:** Worked as a lecturer's project assistant and laboratory assistant coordinators for three courses
- **Organizations** : Himpunan Mahasiswa Rekayasa Pertanian 'Agrapana', Kabinet Mahasiswa ITB (KM-ITB), Skhole - ITB Mengajar

Kyoto University

Master of Agricultural Science
Oct 2021 - Sept 2023

- **Relevant Coursework :** Agricultural Systems Engineering
- **Thesis :** Prediction ofland suitability for soybean production using time series models on weather and land cover data
- **Experiences:** Attended and presented an oral presentation at an international conference on agriculture
- **Organizations** : PPI Kyoto-Shiga

Dibimbing.id

Data Science Bootcamp
Apr 2025 - present

- Learn about fundamental concepts of data science and business understanding
- **Relevant Knowledge :** Data Preprocessing, Database, Data Visualization, Customer Segmentation, Business Understanding with Data Metrics, Prediction Model
- **Technical Skills :** Python, SQL, Tableau, Power BI

EXPERIENCES

Wageningen Food Safety Research

Junior Researcher
Nov 2023 - Aug 2025

- Cleaned and analyzed large datasets of over 300,000 records to assess mycotoxin contamination in European grains and its economic impact
- Performed a comprehensive literature review on economic loss estimation from mycotoxin contamination and created predictive models for crop phenology using remote sensing techniques

---

Climate Change Center ITB

Project Assistant
Dec 2020 - Apr 2021

- Collected and analyzed data on coffee production and climate data across 20+ provinces, identifying key risks to Indonesia's coffee industry
- Conducted a literature review of 10+ studies on climate impacts on the coffee production
- Co-developed a GIS-integrated vulnerability index in collaboration with an agricultural expert and a GIS specialist
- Contributed to project reports and presentations used in stakeholder discussions with the Japanese and Indonesian governments

SKILLS

1 Programming Languages

Python

SQL

3 Visualizations

Tableau

Power BI

Seaborn

Matplotlib

Plotly

2 Frameworks

Pandas

Numpy

Scikit-Learn

SHAP

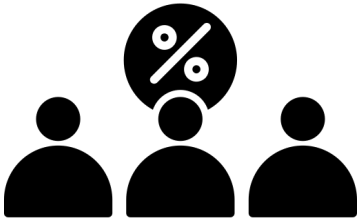
Tensorflow

4 Soft Skills

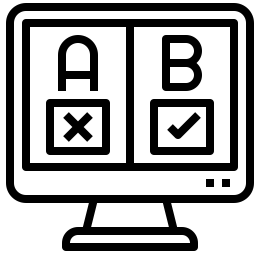
- Communication
- Teamwork
- Data Storytelling
- Analytical Thinking
- Problem-Solving
- Desire to Continue Learning

PROJECTS

E-commerce Transaction Analytics
Analyze sales pattern at e-commerce



Bank Customer Churn Prediction
Develop customer churn prediction model using classification algorithms



A/B Testing on Landing Page Designs
Conduct A/B testing to evaluate the effectiveness of different landing page designs on speaker sales

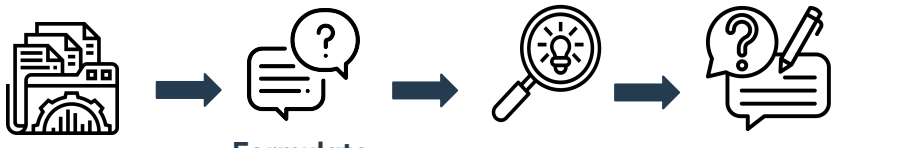
Customer Segmentation of Airline Passengers
Segment airline passengers using K-Means clustering

PROJECT 1 : Data Visualization

Project Description

This project aims to **explore the sales trends of e-commerce** over the period of 2010 and 2011, **gain insights**, and **provide business recommendations** to increase revenue

Methodology



Insights

1. **Most transactions were made by customers during weekdays**, especially Monday - Thursdays afternoons
2. **United Kingdom was the main hub of the e-commerce** due to its highest number of customers and total sales
3. **November become a month with the highest total sales** despite the lowest average order value per transaction which may be happened due to Black Friday
4. **One-time buyer dominated total customers**
5. **Quantity sold was the most influential factor in driving sales instead of price per unit**
6. Netherlands, Ireland, and Austria had exceptionally high average order value despite having low number of customers

Recommendations

1. Give push notifications to increase customer engagement during weekend or weekdays evenings
2. Make a campaign/offer discount during the evening when majority of users are engaged
3. Develop marketing strategies to increase quantity sold
4. Develop loyalty program or redeemed points to increase repeated orders
5. Do market expansion to Netherlands, Ireland, and Austria

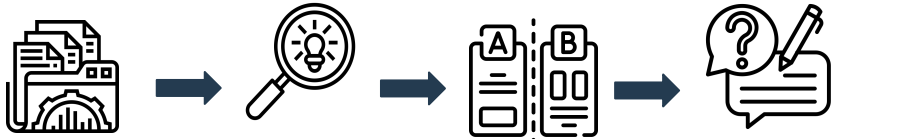
CODE

PROJECT 2: A/B Testing

Project Description

This project involves conducting an **A/B test to compare three landing page designs - Vibrant, Heat, Cold - and determine which design performs better in terms of conversion rates**. Beyond that, **the analysis will consider other factors such as demographic and user behaviours.** The insight will enhance marketing strategies, improve customer experience on the web/app, and boost speaker sales.

Methodology



Data Preprocessing → Exploratory Data Analysis → Hypothesis Testing → Business Recommendations

Results

1. There is **statistically significant difference in conversion rates** between different landing page designs (p value < 0.05) **with cold theme color has the highest conversion rate**
2. **There is no significant difference in conversion rates and average sales across both age group and gender**
3. **Social media has the lowest conversion rate** than other channels even though it is not statistically different when comparing to referral and organic. However, it is proved that **social media has a significantly lower conversion rate (14.45%) vs. Paid (15.68%) (p < 0.05)**

Recommendations

1. Implement cold color theme on the landing page
2. Investigate other factors (button, layout, font color) that can significantly impact customer engagement (page views and session duration) and do another A/B testing
3. Focused on using organic, paid, and referral channels instead of social media
4. Scale up successful campaign on paid channel

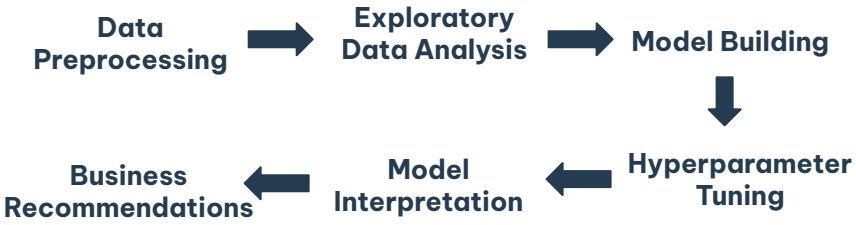
DECK **CODE**

PROJECT 3: Classification Model

Project Description

This project aims to **understand factors influencing customer churn, make prediction model, and give business recommendations** based on the prediction model with bank dataset. This project implements three classification algorithms : Random Forest, Decision Tree, and XGBoost as well as SHAP for feature importance

Methodology



Results

1. XGBoost is the best model with the precision score and recall score on testing data of 0.89
2. **XGBoost can accurately predict 89% of churned customers**
3. The most influential factors on customer churn :
   - Total transactions in the last 12 months
   - Total transaction amounts over the last 12 months
   - Total funds used in one period

Recommendations

1. Improve the quality of bank services by collecting customers' feedbacks
2. Identify and analyze at-risk customers by leveraging prediction model
3. Develop tailored programs for at-risk customers
4. Engage actively with customers to make customers feel valued and appreciated by the bank

CODE

PROJECT 4: Clustering

Project Description

This project aims to segment airline passenger based on LRFMC model which is a modified RFM model. K-Means clustering was chosen for customer segmentation.

Results

There are four customer clusters :

- **Cluster 0 (Lost - New Joiners)**

Passengers recently joined the program, but have not used flights for a long time (436 days); low frequency and monetary

- **Cluster 1 (The Champions)**

Although this cluster ranked second for membership duration, passengers have the highest frequency and monetary

- **Cluster 2 (Need Attention)**

Longest membership duration; low frequency and monetary; have used flights recently

- **Cluster 3 (Recent Passengers - Potential Loyalist)**

Passengers just joined the program more recently, thus, these passengers still have low values of frequency and monetary

Methodology

Recommendations

1. Develop tailored programs for different clusters

CODE

Let’s Connect with Me!

dindararas

Dinda Raraswati

Dinda Raraswati