

Data Science and Business Analytics

Portfolio

By Elizabeth Leonny Efendi

About Me

Elizabeth Leonny Efendi

Enthusiastic and driven Data Science and Business Analytics student from Indonesia, looking to expand my analytical skills and knowledge in a real-world setting. Eager to contribute to a dynamic team, leverage data-related skills, and gain hands on experience.



<https://github.com/gezie1/Portfolio>



Stock Price Prediction: Meta

Language:

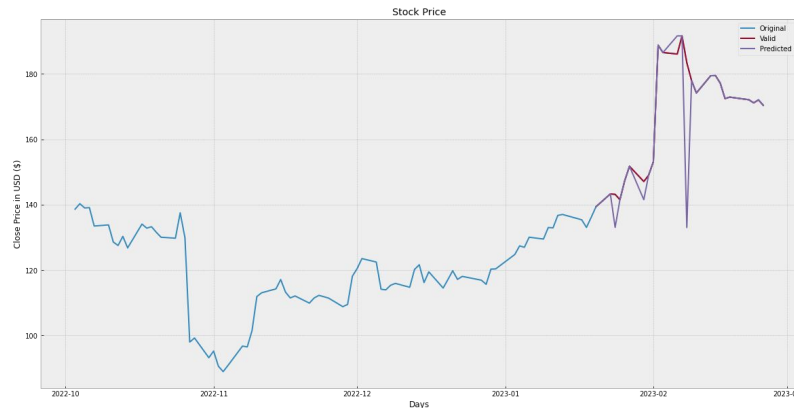
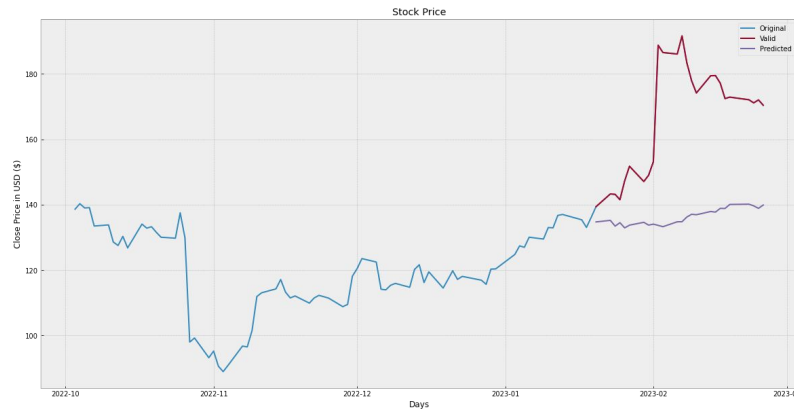


Data:

- Sources: Yahoo Finance
- Features: Moving averages, trading volume, financial ratios

Methodology:

- Data Preprocessing: Handled missing values, feature engineering
- Models: Linear Regression, Decision Tree

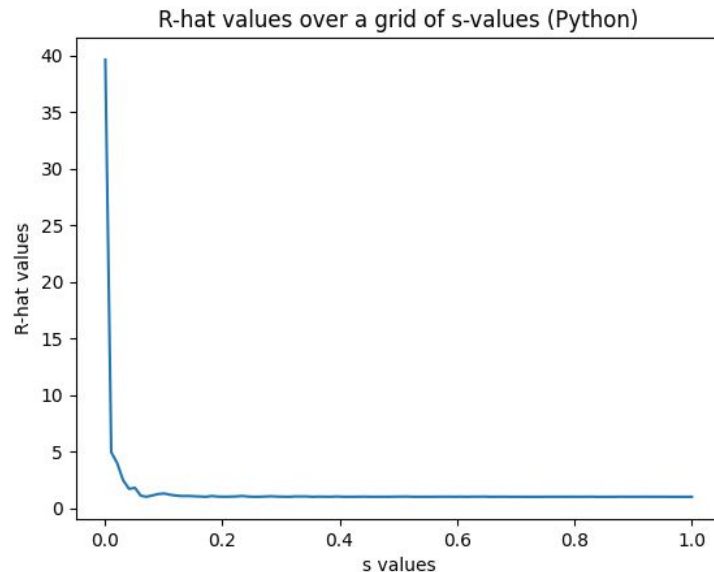
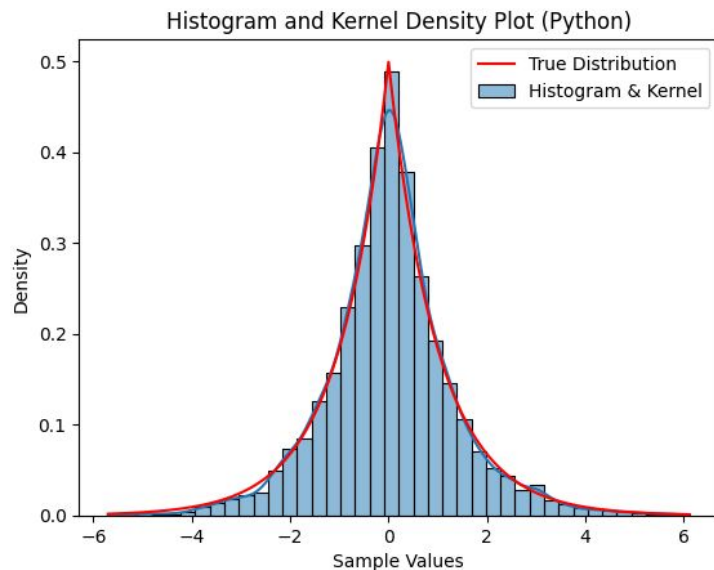


Markov Chain Monte Carlo Simulation

Methodology:

- Models: Metropolis-Hastings Algorithm, Random Walk Metropolis

Language:



Flight Analysis

Language:

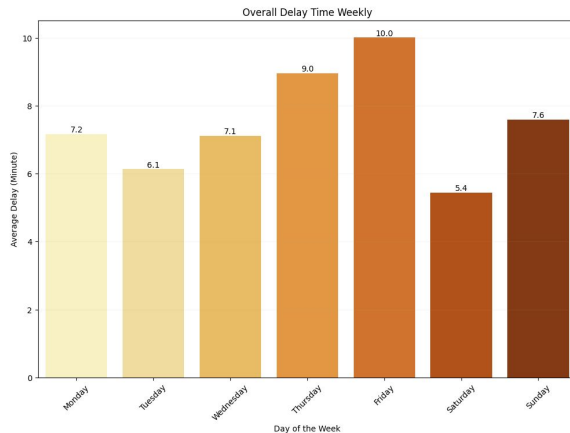
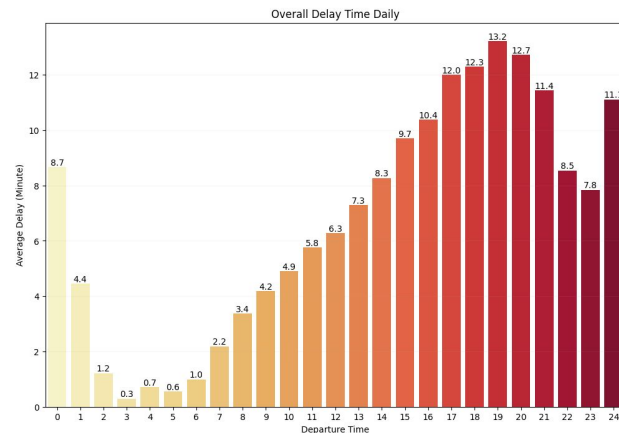


Data:

- Source: The 2009 ASA Statistical Computing and Graphics Data Expo

Methodology:

- Data Collection
- Data Cleaning
- Data Transformation
- Statistical Analysis
- Logistic Regression



Year	Day of the Week	Average Delay
1995	Monday	6.512384
1996	Saturday	7.859997
1997	Tuesday	5.868216
1998	Saturday	5.349508
1999	Saturday	6.780820
2000	Saturday	7.883807
2001	Tuesday	4.779724
2002	Saturday	2.107944
2003	Saturday	2.171316
2004	Saturday	4.215136

Flight Analysis

Language:

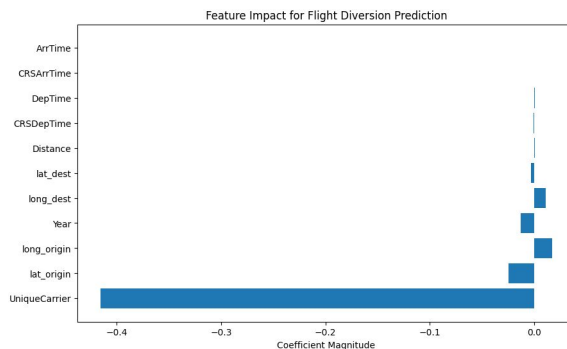
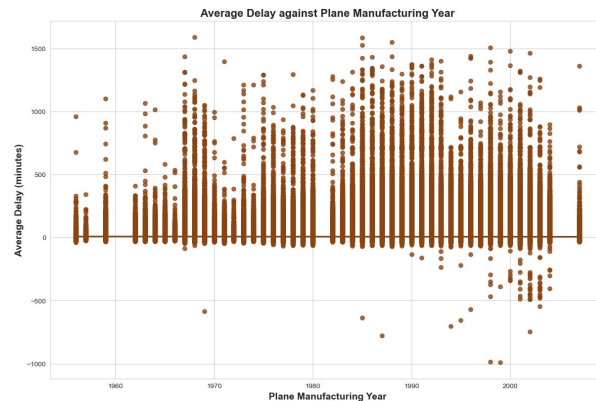


Data:

- Source: The 2009 ASA Statistical Computing and Graphics Data Expo

Methodology:

- Data Collection
- Data Cleaning
- Data Transformation
- Statistical Analysis
- Logistic Regression



Feature coefficients:

UniqueCarrier: -0.41592616010853733
lat_origin: -0.02472788594958321
long_origin: 0.01732631721702439
Year: -0.012916715774805342
long_dest: 0.011310413610952624
lat_dest: -0.003262644191880429
Distance: 0.0006178011191978469
CRSDepTime: -0.00047556415542441486
DepTime: 0.00044985926237693137
CRSArrTime: -8.565315048043337e-05
ArrTime: 5.088578261431095e-05

year	average_delay
1956	8.30874
1957	3.33973
1959	6.89148
1962	6.89828
1963	6.58836
1964	6.59962
1965	4.89256
1966	7.66783
1967	4.91252
1968	5.19571
1969	4.62854
1970	4.47843
1971	3.28762
1972	7.89526
1973	4.83987
1974	5.96586
1975	6.40275
1976	6.83899
1977	6.40152
1978	6.40372
1979	6.3866
1980	6.5621
1982	6.84548
1983	6.54396
1984	8.96743
1985	8.17843
1986	7.81842
1987	7.53597
1988	7.54956
1989	8.82481
1990	8.73876
1991	7.85736
1992	8.40239
1993	7.59530
1994	7.86293
1995	6.70815
1996	7.2654
1997	6.81724
1998	6.22582
1999	5.82449
2000	5.29945
2001	4.79883
2002	5.51111
2003	5.81161
2004	7.79378
2007	5.98738

Flight Analysis

Language:

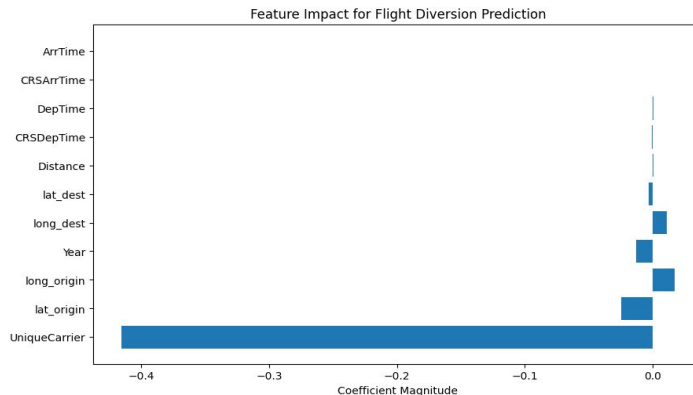
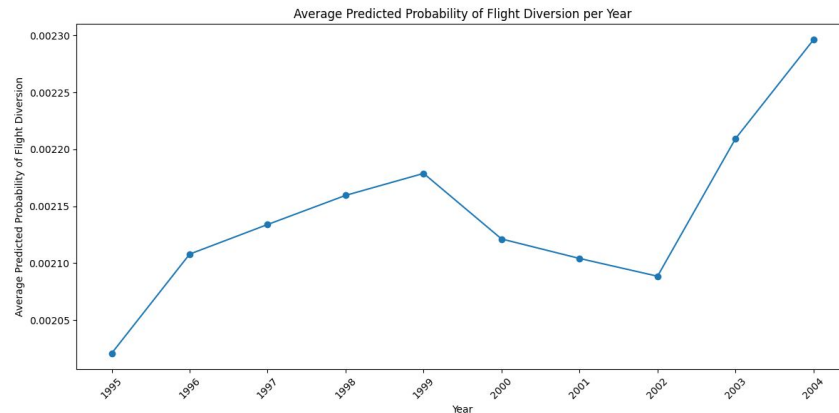


Data:

- Source: The 2009 ASA Statistical Computing and Graphics Data Expo

Methodology:

- Data Collection
- Data Cleaning
- Data Transformation
- Statistical Analysis
- Logistic Regression



Feature coefficients:

```
UniqueCarrier: -0.41592616010853733
lat_origin: -0.02472788594958321
long_origin: 0.01732631721702439
Year: -0.012916715774805342
long_dest: 0.011310413610952624
lat_dest: -0.003262644191880429
Distance: 0.0006178011191978469
CRSDepTime: -0.00047556415542441486
DepTime: 0.00044985926237693137
CRSArrTime: -8.565315048043337e-05
ArrTime: 5.088578261431095e-05
```

Business Analysis: Canadian Market

2.1 Canada Market Overview

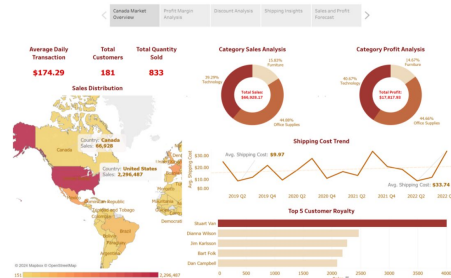


Figure 1: Canada Market Overview

2.2 Profit Margin Analysis



Figure 2: Canada and the US Profit Margin Analysis

2.3 Discount Analysis

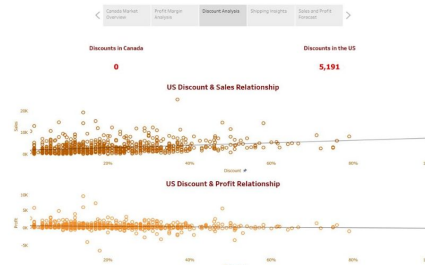


Figure 3: Sales, Discount and Profit Relationship

2.5 Forecasts

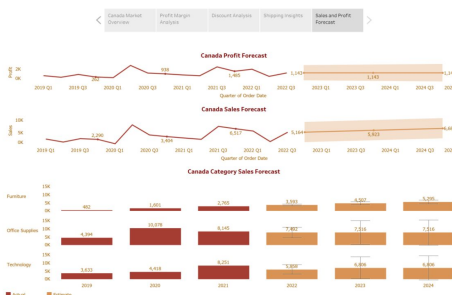


Figure 6: Sales and Profit Forecasts

2.4 Shipping Insights



Figure 4: Shipping Insights 1 (Canada)



Figure 5: Shipping Insights 2 (US)

Research Paper: Clinical Decision Support System

3.1 CDSS Mechanisms

The CDSS processes are complex since it evaluates a large amount of patient data, aligning it with medical literature, case histories, and additional information through advanced algorithms. CDSSs are typically classified into knowledge-based and non-knowledge-based.

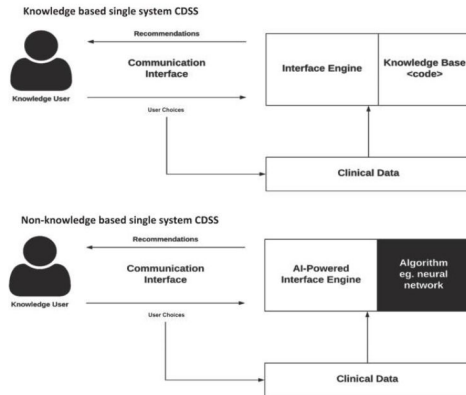


Figure 1: Interactions in knowledge-based and non-knowledge-based CDSS.

Knowledge-based CDSSs employ logical procedures to provide suggestions to help physicians. There will always be knowledge sources and rules obtained from medical literature, patient-centered procedures, guidelines, and expert knowledge. It is commonly used to handle complex decision-making cases. Meanwhile, non-knowledge based CDSSs use complex algorithms to make medical decisions and are usually used when a medical case has not explicitly happened in any past scenarios. One of the algorithms used is neural networks, which is a branch of AI to teach computers the way the human brain processes information and learns.

3.3.3 Reliability of Data

The reliability of data is crucial for providing high-quality patient care and making informed decisions. Some research found that patient data from EHRs are not entirely accurate, which is likely due to a lack of EHR usability (Dash et al., 2019). However, maintaining data reliability is a significant challenge. To input and update data manually requires time and energy, and is prone to errors, leading to inaccurate and incomplete data.

Conducting routine data audits can assist in verifying the precision and accuracy of the data. Also, healthcare providers might use automated data input and update processes to reduce errors and save time and energy. This includes using data import and export tools, integrating data systems, and implementing automated workflows.

	Mean Completeness Score	Mean Correctness Score
Hip Pain	.39	.91
Shoulder Pain	.32	.94
Knee Pain	.37	.96
Foot Pain	.30	.95
All cases combined	.34	.94

Figure 7: Completeness and Correctness Scores

Research conducted by internal medicine residents (PGY-1-3), shows that the core issue is completeness of data. Of the six elements, the data entered is only 30%-40% of the total data that should be entered. However, the data's average accuracy rate of 94%.

5.1 Strategic Planning

Strategic planning plays a crucial role in outlining the course of an organization and deciding how to allocate its resources to pursue that path (Reynolds, G.W., 2016). Implementation of CDSS can be classified as a growth or innovation project, which generates significant new revenue for the organization while exploring the use of new technology in a new way at the same time. With strategic planning, healthcare organizations can align their CDSS initiatives with their overall business objectives such as enhancing customer satisfaction and defining pricing strategies to sustain a competitive edge. Analyzing user feedback and satisfaction ratings can reveal areas for improvement in CDSS offerings. Furthermore, the CDSS market is expected to experience substantial growth in the future. Healthcare organizations can analyze data on the adoption rates of CDSS to inform pricing strategies.

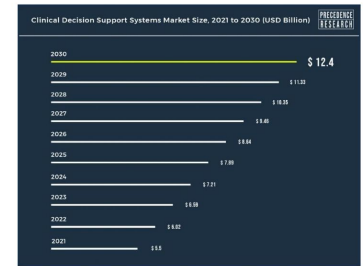


Figure 8: CDSS Market Size Prediction

Stay In Touch



elizabethleonnnyefendi@gmail.com



<https://github.com/gezie1/Portfolio>