# UNIVERSITI KUALA LUMPUR
## ASSESSMENT BRIEF

| COURSE DETAILS | |
|---|---|
| **INSTITUTE** | UniKL BRITISH MALAYSIAN INSTITUTE |
| **COURSE NAME** | BIG DATA ANALYTICS |
| **COURSE CODE** | BEB43403 |
| **COURSE LEADER** | MUHD KHAIRULZAMAN ABDUL KADIR |
| **LECTURER** | MUHD KHAIRULZAMAN ABDUL KADIR |
| **SEMESTER & YEAR** | OCTOBER 2023 |

| ASSESSMENT DETAILS | |
|---|---|
| **TITLE/NAME** | CASE STUDY |
| **WEIGHTING** | 25% |
| **DATE/DEADLINE** | 19/12/2023, 5.00PM |
| **COURSE LEARNING OUTCOME(S)** | CLO3: Perform analysis of various data driven predictive models for effective business decisions (A5, PLO12). |
| **INSTRUCTIONS** | Perform the following tasks:<br>1. Read the instructions given in the assessment sheet CAREFULLY.<br>2. Answer all tasks.<br>3. Answers must be in English. |

| **Student Name:** | **ID:** | **Group:** |
|---|---|---|
| **1) Anis Suriani binti Azizan** | **(51221221056)** | |
| **2) Nurul Izzati binti Mahamad Ismail** | **(51221221147)** | **Group 3** |
| **3) Ahmad Syahmi bin Ahmad Pauzi** | **(51221221003)** | **(L01)** |
| **4) Muhammad Waiz bin Nor Kamal** | **(51221221053)** | |
| **5) Muhammad Ikhwan Syafiq bin Norsham** | **(51221221125)** | |
| **Assessor's Comment:** | | **Marks:** |

| **Verified by: Course Leader [MUZA]**<br>**Prepared by: [MUZA]** | **QSC format verification** | **PC/HOS content validation** |
|---|---|---|
| I hereby declare that all my team members have agreed with this assessment. All team members are certain that this assessment complies with the Course Syllabus.<br><br>Signature: _____<br><br>Date : 23/11/2023 | | |

| TASK NO | CLO | MARKING SCHEME | MARKS |
|---|---|---|---|
| 1 | 3 | **Propose** one application used big data analytics in any industry areas. | 15 |
| 2 | 3 | **Show** and **explain** how the big data is used in the proposed area of application (Detail process – can be either in a block diagram or flowchart or any other method). | 15 |
| 3 | 3 | **Give** and **show** the demo or coding being generated for the big data application proposed. | 15 |
| 4 | 3 | **List** the **benefits and challenges** (at least 5 each) of implementing big data analytics to the area proposed to industry and society. | 10 |
| 5 | 3 | **Recommend** the best platform for big data analytics you can use and **explain** why the platform you selected is the best. (It can be either related to the proposed application or not related to it). | 15 |
| 6 | 3 | **Suggest TWO (2)** big data analytics algorithm or techniques and **explain** how these 2 techniques work. | 10 |
| 7 | 3 | **Give** TWO (2) or more examples of other than the application proposed in any area that used big data analytics. | 10 |
| 8 | 3 | **Present** all the ideas clearly and effectively with Q&A from both parties (Team presenter & member of the floor) | 10 |
| | | **TOTAL** | **100** |

# INFORMATION ON SK_SP-TA FOR COURSE

| | | |
|---|---|---|
| **Course Code & Name** | : | BEB43403 **& BIG DATA ANALYTICS** |
| **PLOs** | : | **12** |

**Please tick ( ) in the box provided.**

| ***Knowledge Profiles (SK)*** A programme that builds this type of knowledge and develops the attributes listed below is typically achieved in 4 years of study | | |
|---|---|---|
| **SK1** | A systematic, theory-based understanding of the natural sciences applicable to the sub-discipline | |
| **SK2** | Conceptually-based mathematics, numerical analysis, statistics and aspects of computer and information science to support analysis and use of models applicable to the sub-discipline | |
| **SK3** | A systematic , theory-based formulation of engineering fundamentals required in an accepted sub-discipline | |
| **SK4** | Engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for an accepted sub-discipline | |
| **SK5** | Knowledge that supports engineering design using the technologies of a practice area | |
| **SK6** | Knowledge of engineering technologies applicable in the sub-discipline | |
| **SK7** | Comprehension of the role of technology in society and identified issues in applying engineering technology: ethics and impacts: economic, social, environmental and sustainability | |
| **SK8** | Engagement with the technological literature of the discipline | |

| ***Definition of Broadly-Defined Problem Solving (SP)*** | | | |
|---|---|---|---|
| **No.** | **Attribute** | **Broadly-defined Engineering Problems have characteristic SP1 and some or all of SP2 to SP7:** | |
| **SP1** | **Depth of Knowledge Required** | Cannot be resolved without engineering knowledge at the level of one or more of SK 4, SK5, and SK6 supported by SK3 with a strong emphasis on the application of developed technology | |
| **SP2** | **Range of conflicting requirements** | Involve a variety of factors which may impose conflicting constraints. | |
| **SP3** | **Depth of analysis required** | Can be solved by application of well-proven analysis techniques | |
| **SP4** | **Familiarity of issues** | Belong to families of familiar problems which are solved in well-accepted ways | |
| **SP5** | **Extent of applicable codes** | May be partially outside those encompassed by standards or codes of practice | |
| **SP6** | **Extent of stakeholder involvement and level of conflicting requirements** | Involve several groups of stakeholders with differing and occasionally conflicting needs | |
| **SP7** | **Interdependence** | Are parts of, or systems within complex engineering problems | |

| ***Range of Engineering Activities (TA)*** | | | |
|---|---|---|---|
| **No.** | **Attribute** | **Broadly-defined activities** | |
| **TA1** | **Range of resources** | Involve a variety of resources (and for this purposes resources includes people, money, equipment, materials, information and technologies) | |
| **TA2** | **Level of interactions** | Require resolution of occasional interactions between technical, engineering and other issues, of which few are conflicting | |
| **TA3** | **Innovation** | Involve the use of new materials, techniques or processes in non-standard ways | |
| **TA4** | **Consequences to society and the environment** | Have reasonably predictable consequences that are most important locally, but may extend more widely | |
| **TA5** | **Familiarity** | Require a knowledge of normal operating procedures and processes | |

## CASE STUDY

Big data is reshaping various industries, especially during the transition from Industry Revolution 3 to Industry Revolution 4. Recently, big data has advanced from conventional technology to sophisticated tools like cloud computing and artificial intelligence. It is now extensively employed in sectors like finance, healthcare, and manufacturing to enhance processes, boost efficiency, and gain fresh insights.

For instance, banks utilize big data to detect fraud and evaluate credit risk, while hospitals analyze patient data to enhance treatment outcomes. Nonetheless, working with big data comes with challenges. Maintaining data privacy and security is crucial, and ethical considerations are vital when using data to make decisions impacting people's lives. Despite these hurdles, big data holds significant potential for fostering innovation and shaping the future of industry.

1. Prepare a presentation slide which consists of below information: -.
    - **Propose** one application used big data analytics in any industry areas.
    - **Show** and **explain** on how the big data is used in the proposed area of application (Detail process – can be either in a block diagram or flowchart or any other method) (SP1)
    - **Give** and **show** the demo or coding being generated for the big data application proposed. (SP1)
    - **List** the **benefits and challenges** (at least 5 each) of implementing big data analytics to the area proposed to industry and society.
    - **Recommend** the best platform for big data analytics you can use and **explain** why the platform you selected is the best. (It can be either related to the proposed application or not related to it). (SP2)
    - **Suggest TWO (2)** big data analytics algorithms or techniques and **explain** how these 2 techniques work. (SP3)
    - **Give** TWO (2) or more examples of other than the application proposed in any area that used big data analytics.

2. Maximum duration of presentation is **15 – 20 minutes**.
3. Maximum team size: **5 people**. Slides must clearly list the name and student ID of team members.
4. **All team members** must participate in the case study (name and student ID must be stated prior to presentation).
5. **Please also show and give demo on any coding involved for your study case.**

**CASE STUDY RUBRIC**

**Group member:**

| No | Criteria on Case Study Presentation | Null | Very Poor | Poor | Slightly Insufficient | Sufficient | Exceed Expectation |
|---|---|---|---|---|---|---|---|
| 1.0 | **Propose** one application used big data analytics in any industry areas. | **0** | **1** | **4** | **7** | **10** | **15** |
| 2.0 | **Show** and **explain** on how the big data is used in the proposed area of application (Detail process – can be either in a block diagram or flowchart or any other method) | **0** | **1** | **4** | **7** | **10** | **15** |
| 3.0 | **Give** and **show** the demo or coding being generated for the big data application proposed. | **0** | **1** | **4** | **7** | **10** | **15** |
| 4.0 | **List** the **benefits and challenges** (at least 5 each) of implementing big data analytics to the area proposed to industry and society. | **0** | **2** | **4** | **6** | **8** | **10** |
| 5.0 | **Recommend** the best platform for big data analytics you can use and **explain** why the platform you selected is the best. (It can be either related to the proposed application or not related to it). | **0** | **1** | **4** | **7** | **10** | **15** |
| 6.0 | **Suggest TWO (2)** big data analytics algorithm or techniques and **explain** how these 2 techniques work. | **0** | **2** | **4** | **6** | **8** | **10** |
| 7.0 | **Give** TWO (2) or more examples of other than the application proposed in any area that used big data analytics. | **0** | **2** | **4** | **6** | **8** | **10** |
| 8.0 | **Present** all the ideas clearly and effectively with Q&A from both parties (Team presenter & member of the floor) | **0** | **2** | **4** | **6** | **8** | **10** |
| | **Total case study results** | **/100** | | | | | |

**Assessor comments:**

.

BEB43403 BIG DATA ANALYTICS

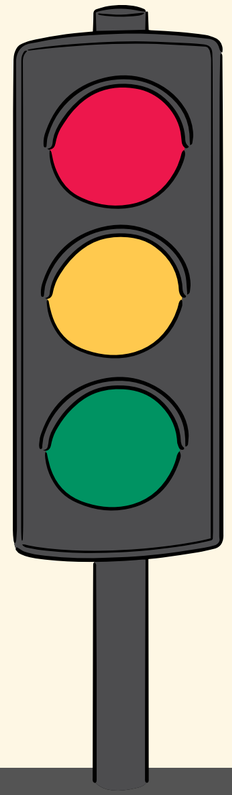
# Case Study : Transportation


Presenting By :

1) Anis Suriani binti Azizan (51221221056)
2) Nurul Izzati binti Mahamad Ismail (51221221147)
3) Ahmad Syahmi bin Ahmad Pauzi (51221221003)
4) Muhammad Waiz bin Nor Kamal (51221221053)
5) Muhammad Ikhwan Syafiq bin Norsham (51221221125)


# Contents





Anis Suriani binti Azizan (51221221056)

# Big Data Analytics in Transportation Area

In transportation, Big Data Analytics is the game-changer. It's the tool that helps us navigate traffic smarter, predict when vehicles need fixing, and make our journeys safer and more efficient.

Anis Suriani binti Azizan (51221221056)

# Block Diagram for Traffic Analytics

## REAL TIME DATA

- GPS sensors (vehicles)
- Traffic cameras
- Tollbooth data

## DATA STORAGE

- Streaming Data
- Batch Data

## TRAFFIC PREDICTION

- Historical and real-time data analysis
- Forecasting traffic patterns and congestion
- Machine learning models

## ALERTS

- Notifications for critical events or anomaliesn

## MODEL UPDATES

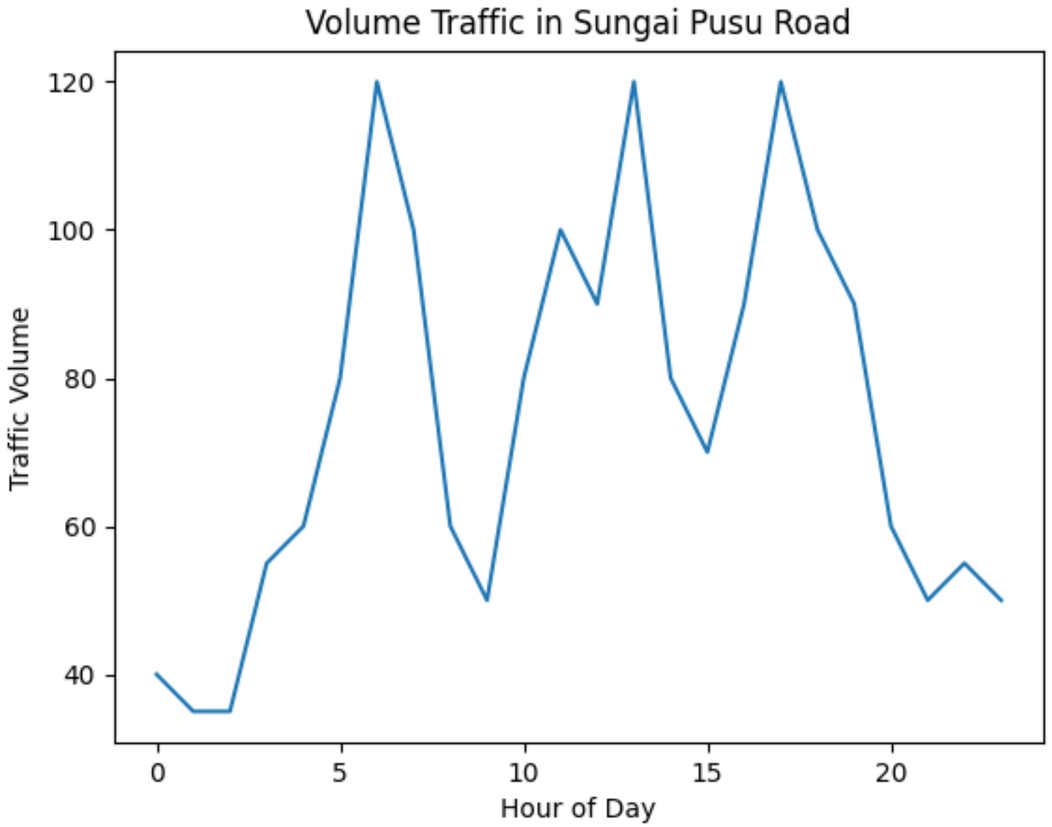
- Retraining models with new data
- Enhancing accuracy and performance

Nurul Izzati binti Mahamad Ismail (51221221147)

# Code for Traffic Pattern Analysis in Sungai Pusu, Gombak

## Data Traffic

| hour | traffic_volume |
|---|---|
| 0 | 40 |
| 1 | 35 |
| 2 | 35 |
| 3 | 55 |
| 4 | 60 |
| 5 | 80 |
| 6 | 120 |
| 7 | 100 |
| 8 | 60 |
| 9 | 50 |
| 10 | 80 |
| 11 | 100 |
| 12 | 90 |
| 13 | 120 |
| 14 | 80 |
| 15 | 70 |
| 16 | 90 |
| 17 | 120 |
| 18 | 100 |
| 19 | 90 |
| 20 | 60 |
| 21 | 50 |
| 22 | 55 |
| 23 | 50 |



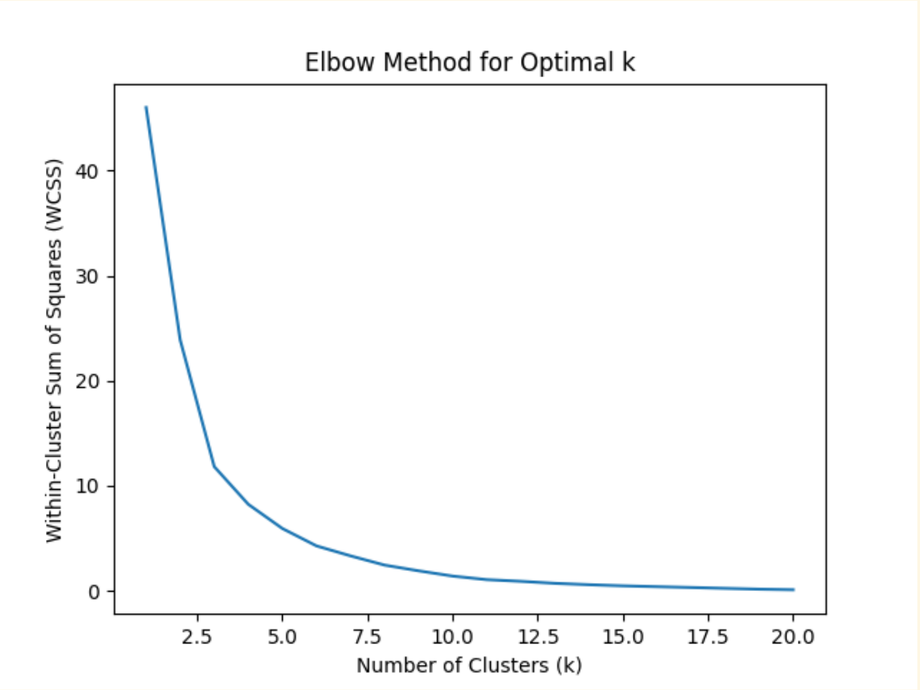

## Graph before Clustering

Nurul Izzati binti Mahamad Ismail (51221221147)

# Result

```python
import pandas as pd
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

# Assuming you have a dataset with traffic-related features (e.g., time of day, traffic volume)
# Replace 'your_dataset.csv' with the actual path or URL of your dataset
data = pd.read_csv('traffic_flow.csv')

# Assuming 'hour' and 'traffic_volume' are relevant features for analysis
features = data[['hour', 'traffic_volume']]

# Standardize the features (important for k-means)
features_standardized = (features - features.mean()) / features.std()

# Determine the optimal number of clusters (k) using the elbow method
wcss = []
for i in range(1, 21):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
    kmeans.fit(features_standardized)
    wcss.append(kmeans.inertia_)

# Plot the elbow curve
plt.plot(range(1, 21), wcss)
plt.title('Elbow Method for Optimal k')
plt.xlabel('Number of Clusters (k)')
plt.ylabel('Within-Cluster Sum of Squares (WCSS)')
plt.show()
```

Code

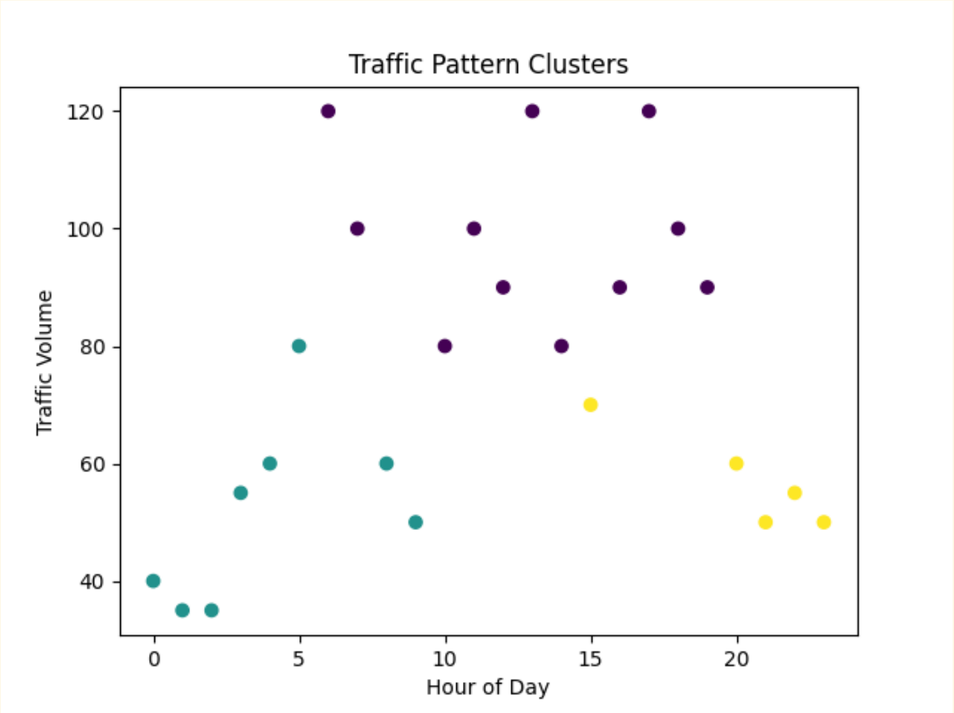


Nurul Izzati binti Mahamad Ismail (51221221147)

# Result

```python
# Based on the elbow method, choose an optimal value for k
optimal_k = 3

# Apply k-means clustering with the chosen k
kmeans = KMeans(n_clusters=optimal_k, init='k-means++', max_iter=300, n_init=10, random_state=0)
features['cluster'] = kmeans.fit_predict(features_standardized)

# Visualize the clusters
plt.scatter(features['hour'], features['traffic_volume'], c=features['cluster'], cmap='viridis')
plt.title('Traffic Pattern Clusters')
plt.xlabel('Hour of Day')
plt.ylabel('Traffic Volume')
plt.show()
```

Code




Graph After Clustering

Ahmad Syahmi bin Ahmad Pauzi (51221221003)

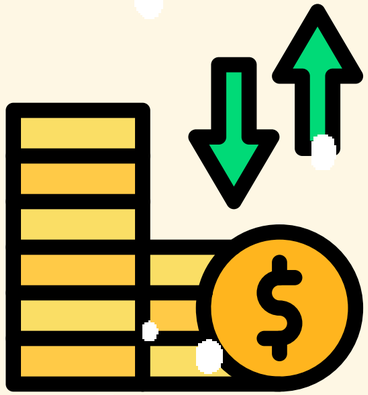
# Benefits

## Improved Customer Experience

## Environmental Sustainability

## Optimized Traffic Management


## Cost Reduction and Resource Optimization



## Innovations in Autonomous Vehicles

Ahmad Syahmi bin Ahmad Pauzi (51221221003)

# Challenges

1. Data privacy and security
2. Cost and infrastructure requirements
3. Data quality and integration issues
4. Public acceptance and trust
5. Technological limitations

# Platform for Big Data Analytics

Muhammad Ikhwan Syafiq bin Norsham (51221221125)

## Apache Kafka

- Real-time Data Processing
- Scalability
- Event-Driven Architecture
- Integration

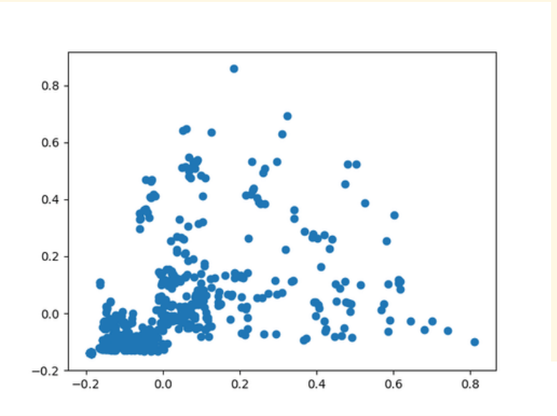
Muhammad Ikhwan Syafiq bin Norsham (51221221125)

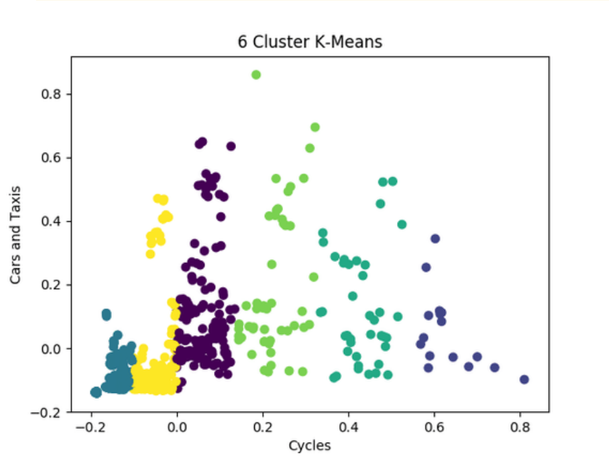
# Big Data Analytics Algorithm for Transportation Area

## K-MEANS CLUSTERING

How it works:

- Data Preparation: Traffic data is collected from various sources such as GPS sensors, cameras, and toll booths.
- Feature Engineering: Relevant features are extracted from the data, such as average speed, travel time, and distance traveled.
- Cluster Identification: The K-Means algorithm iteratively groups data points into k clusters based on their feature similarities.
- Cluster Interpretation: Each cluster is analyzed to understand its characteristics.



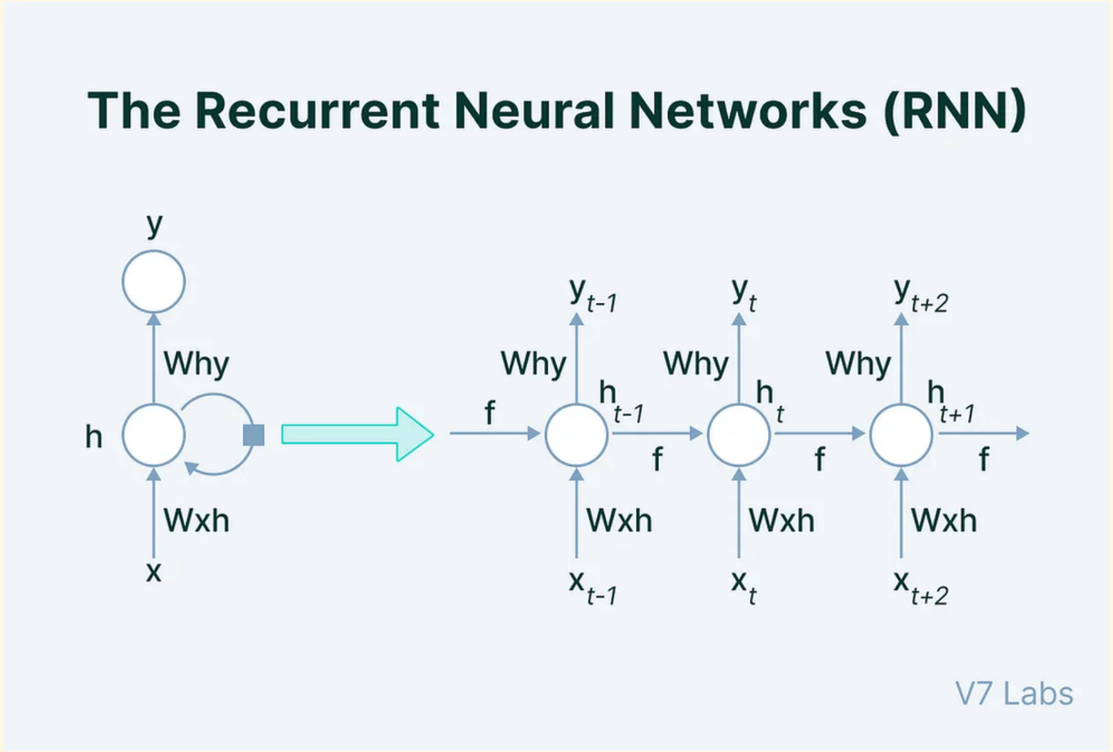
Muhammad Waiz bin Nor Kamal (51221221053)

# Big Data Analytics Algorithm for Transportation Area

## Recurrent Neural Networks (RNNs) for Traffic Prediction

How it works:

- Input Data :Historical traffic data including speed, number of vehicles, and weather conditions.
- Model Training: The RNN model takes sequences of data points as input and learns to predict the next value in the sequence.
- Real-time Updates: The trained model can be used with real-time traffic data from sensors and cameras to continuously update its predictions and provide accurate forecasts for short-term (minutes to hours) traffic conditions.



# Example of Big Data Application in Industry Area

Muhammad Waiz bin Nor Kamal (51221221053)

## HEALTHCARE

- Personalized Treatment
- Remote Patient Monitoring
- Clinical Research

## MANUFACTURING

- Predictive Maintenance
- Supply Chain Optimization.
- Quality Control

## RETAIL

- Personalized Recommendation
- Inventory Management
- Fraud Detection

## FINANCIAL SERVICES

- Risk Management
- Customer Analytics
- Algorithmic Trading

## TRANSPORTATION

- Traffic Optimization
- Route Planning
- Autonomous Vehicles

## ENERGY

- Smart Grids
- Renewable Energy
- Demand Forecasting

# Thank you for listening!

Don't hesitate to ask any questions!