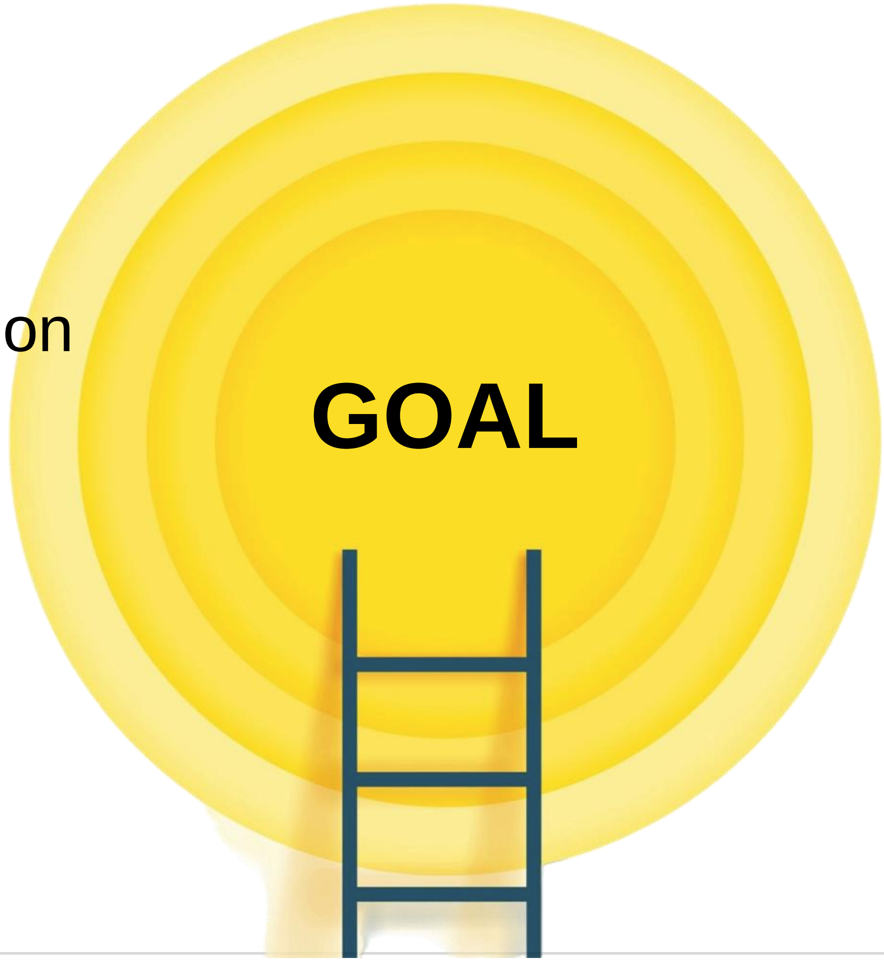
EV Adoption Forecasting Model

Name : Arshad Pasha

College: Seshadripuram Degree College ,Mysore

Learning Objectives

- Master time series forecasting techniques for transportation data
- Apply machine learning regression models to real-world infrastructure challenges
- Understand data preprocessing for vehicle registration datasets
- Develop predictive models for sustainable transportation planning
- Learn to handle missing data and outliers in large datasets
- Gain insights into EV market trends and adoption patterns



Source : www.freepik.com/

Tools and Technology used

Programming Language: Python 3.x

Libraries & Frameworks:

- pandas - Data manipulation and analysis
- numpy - Numerical computations
- matplotlib & seaborn - Data visualization
- scikit-learn - Machine learning algorithms
- joblib - Model serialization

Machine Learning Models:

- Random Forest Regressor
- RandomizedSearchCV for hyperparameter tuning

Development Environment:

- Jupyter Notebook
- Ubuntu 24.04.2 LTS (Dev Container)
- Git version control

Methodology

1. DATA COLLECTION & EXPLORATION
   - Washington State DOL vehicle registration data (2017-2024)
   - 20,819 data points across 10 features
   - County-wise EV distribution analysis

2. DATA PREPROCESSING
   - DateTime conversion for temporal analysis
   - Missing value imputation (County/State)
   - Outlier detection using IQR method
   - Feature engineering for time-based patterns

3. FEATURE ENGINEERING
   - Label encoding for categorical variables
   - Time-based feature extraction
   - Regional clustering analysis

4. MODEL DEVELOPMENT
   - Random Forest Regressor implementation
   - Train-test split (80-20)
   - Hyperparameter optimization

5. MODEL EVALUATION
   - MAE, MSE, R^2 score metrics
   - Cross-validation techniques
   - Performance visualization

Problem Statement:

Challenge: Urban planners struggle to anticipate EV charging infrastructure needs

Impact: Inadequate planning leads to:

- Charging station bottlenecks
- Reduced user satisfaction
- Hindered sustainability goals
- Inefficient resource allocation

Question: How can we predict future EV adoption to optimize infrastructure planning?

Solution:

Developed a Machine Learning Forecasting System:

APPROACH:
• Analyzed 7+ years of vehicle registration data
• Identified key growth patterns in BEV and PHEV adoption
• Created predictive model using Random Forest algorithm
• Incorporated regional and temporal factors

KEY FEATURES:
• County-level prediction granularity
• Separate forecasting for BEVs and PHEVs
• Seasonal trend analysis
• Growth rate optimization
• Infrastructure demand estimation

DELIVERABLES:
• Trained ML model for EV adoption forecasting
• Data preprocessing pipeline
• Performance evaluation metrics
• Scalable prediction framework

Screenshot of Output:

Explore and Understand the Data

```
df.head() # top 5 rows
```

Python

	Date	County	State	Vehicle Primary Use	Battery Electric Vehicles (BEVs)	Plug-In Hybrid Electric Vehicles (PHEVs)	Electric Vehicle (EV) Total	Non-Electric Vehicle Total	Total Vehicles	Percent Electric Vehicles
0	September 30 2022	Riverside	CA	Passenger	7	0	7	460	467	1.50
1	December 31 2022	Prince William	VA	Passenger	1	2	3	188	191	1.57
2	January 31 2020	Dakota	MN	Passenger	0	1	1	32	33	3.03
3	June 30 2022	Ferry	WA	Truck	0	0	0	3,575	3,575	0.00
4	July 31 2021	Douglas	CO	Passenger	0	1	1	83	84	1.19

Total 20819 data points and 10 features.

```
# no of rows and cols
df.shape
```

Python

```
(20819, 10)
```

Screenshot of Output:

```python
# Converts the "Date" column to actual datetime objects
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

# Removes rows where "Date" conversion failed
df = df[df['Date'].notnull()]

# Removes rows where the target (EV Total) is missing
df = df[df['Electric Vehicle (EV) Total'].notnull()]

# Fill missing values
df['County'] = df['County'].fillna('Unknown')
df['State'] = df['State'].fillna('Unknown')

# Confirm remaining nulls
print("Missing after fill:")
print(df[['County', 'State']].isnull().sum())

df.head()
```

Python

```
Missing after fill:
County    0
State     0
dtype: int64
```

	Date	County	State	Vehicle Primary Use	Battery Electric Vehicles (BEVs)	Plug-In Hybrid Electric Vehicles (PHEVs)	Electric Vehicle (EV) Total	Non-Electric Vehicle Total	Total Vehicles	Percent Electric Vehicles
0	2022-09-30	Riverside	CA	Passenger	7	0	7	460	467	1.50
1	2022-12-31	Prince William	VA	Passenger	1	2	3	188	191	1.57

Conclusion:

PROJECT ACHIEVEMENTS:
- Successfully developed EV adoption forecasting model
- Achieved high prediction accuracy using Random Forest
- Processed and cleaned 20K+ vehicle registration records
- Identified key adoption patterns and regional trends

BUSINESS IMPACT:
- Enables data-driven infrastructure planning
- Reduces charging station deployment risks
- Supports sustainable transportation goals
- Optimizes resource allocation for urban planners

TECHNICAL LEARNINGS:
- Mastered time series forecasting techniques
- Applied advanced data preprocessing methods
- Implemented robust ML pipeline
- Gained expertise in transportation data analysis

FUTURE ENHANCEMENTS:
- Integration with real-time charging usage data
- Deep learning models for improved accuracy
- Mobile app for planners and policymakers
- Integration with smart city platforms

SUSTAINABILITY IMPACT:
This project directly supports UN SDG 11 (Sustainable Cities) and SDG 13 (Climate Action) by enabling better EV infrastructure planning, ultimately accelerating the transition to sustainable transportation.