Truck Activity Pattern Classification Using Anonymous Mobile Sensor Data

TRB Innovations in Freight Data Workshop

Taslima Akter
Doctoral Candidate, Department of Civil Engineering
University of Arkansas

Advisor: Dr. Sarah Hernandez
Assistant Professor, Department of Civil Engineering
University of Arkansas

Problem Statement

- Truck GPS data, a valuable source of freight movement
- Large data stream, how to deduce insights
- Does not contain industry type
- Difficult to use in commodity based freight forecasts

Research Question

Can
truck activity patterns
by
industry type
be discerned from truck GPS data while maintaining anonymity?

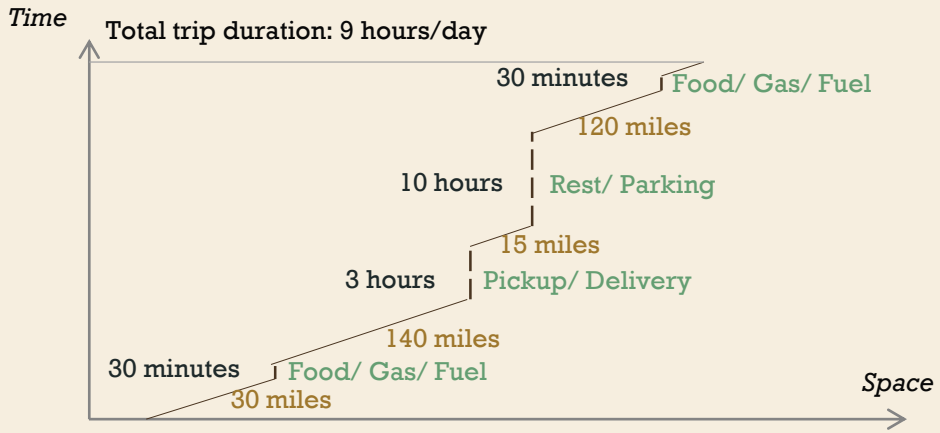
Applications

Activity Profiles for Activity Based Models (ABM)

Commodity-based Freight Travel Demand Models

Methods "Teaser"

Activity Pattern Profiles
Six common patterns found in AR



Support Activity Based Models for Freight using GPS Data to derive **activity patterns**

Industry Classification
Eight industry groups classified by activity pattern

Characteristics of:

1. Trip
2. Stops
3. Land Uses

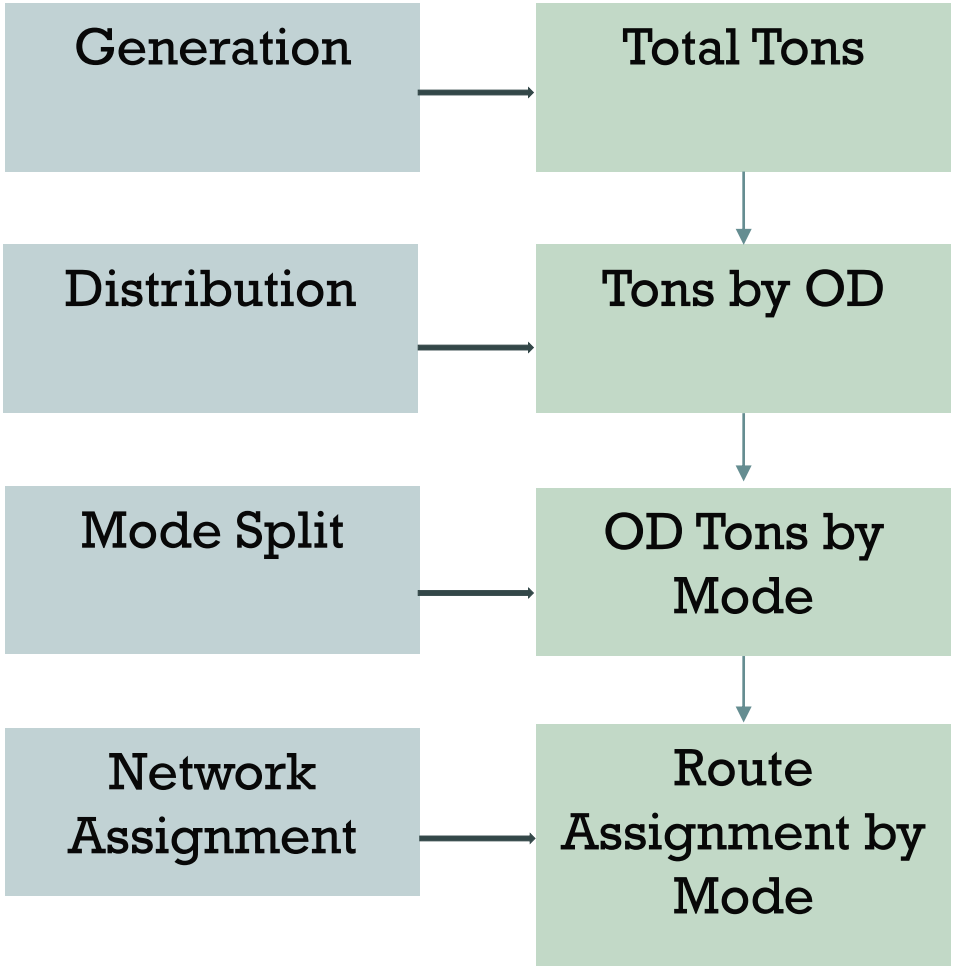
- Agriculture/Livestock
- Oil and Gas/ Quarry
- Consumer Products
- Machinery Equipment/ Chemicals
- Pass-through

Support Commodity-based Freight Travel Demand Models using GPS Data to derive **industry classification**

Outline

1. Background and Motivations
   A. Commodity Based Freight Models
   B. Activity Based Freight Models
2. Activity Pattern Classification Method and Application
3. Industry Classification Method and Application
4. Future Research

Overview of Commodity Based Models

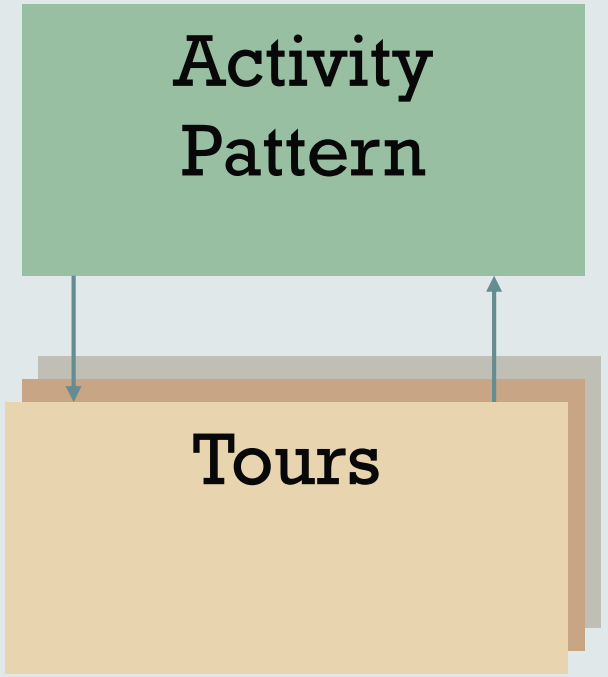


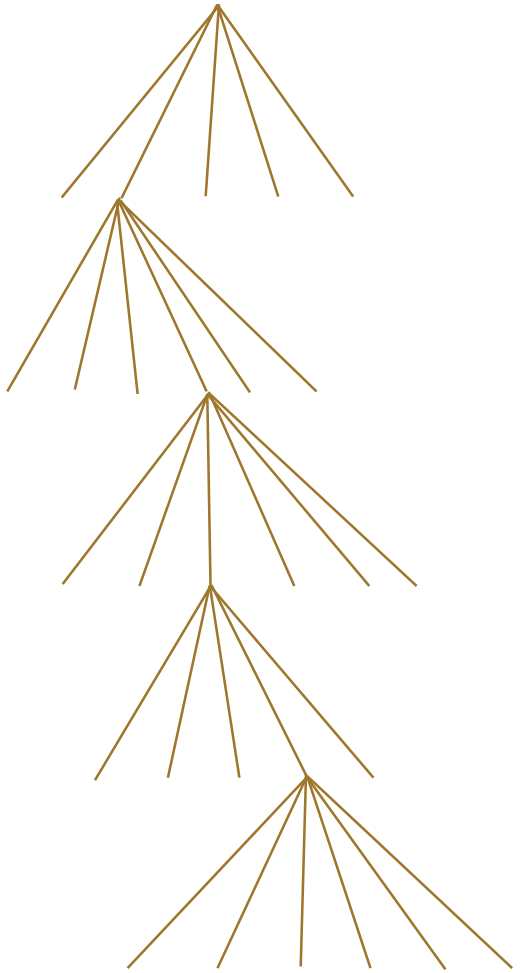
Commodity Group	Commodity Name
1	Agriculture
2	Mining
3	Coal
4	Nonmetallic Minerals
5	Food
6	Consumer Manufacturing
7	Non-Durable Manufacturing
8	Lumber
9	Durable Manufacturing
10	Paper
11	Chemicals
12	Petroleum
13	Clay, Concrete, Glass
14	Primary Metal
15	Secondary & Misc. Mixed

Source: Arkansas Statewide Travel Demand Model, 2015

Overview of Activity Based Models

Activity Schedule





- Activity Pattern
- Primary Tour Time of Day
- Primary Tour Destination and Mode
- Secondary Tour Time of Day
- Secondary Tour Destination and Mode

Source: Bowman and Ben-Akiva, 2001

Comparison of Different Models

	Four-Step Model	Activity Based Model
Advantages	▪ Predicts the aggregated flows between zones ▪ Historically popular models ▪ Ease of development	▪ A disaggregated model ▪ Predicts the detailed flows considering: • Behavioral aspects • Individual operational decisions • Interactions between supply chain components
Limitations	▪ Aggregation may limit prediction accuracy	▪ Lack of available data ▪ Lots of agents increase the complexity of the model

Activity Profile Generation for Passenger Vehicle

"ABM system represents a person's choice of activities and associated travel as an activity pattern overarching a set of tours."

Source: Bowman and Ben-Akiva, 2001

[1] Features

- Socio-demographic attributes
- Household data
- Vehicle data
- Land use data
- Travel events
- Travel behavior
- Time of day
- Mode choice

[2] Data Source

- Regional planning studies
- Census data
- Surveys (e.g., household survey, travel survey, activity diaries, GPS-based travel survey, etc.)
- Social media (e.g., Twitter)

[1, 2] Source: *Recker, 2001; Allahviranloo and Recker, 2013; Hasan and Ukkusuri, 2014; Chung and Shalaby, 2005*

Activity Profile Generation for Freight

[1] Feature Extraction

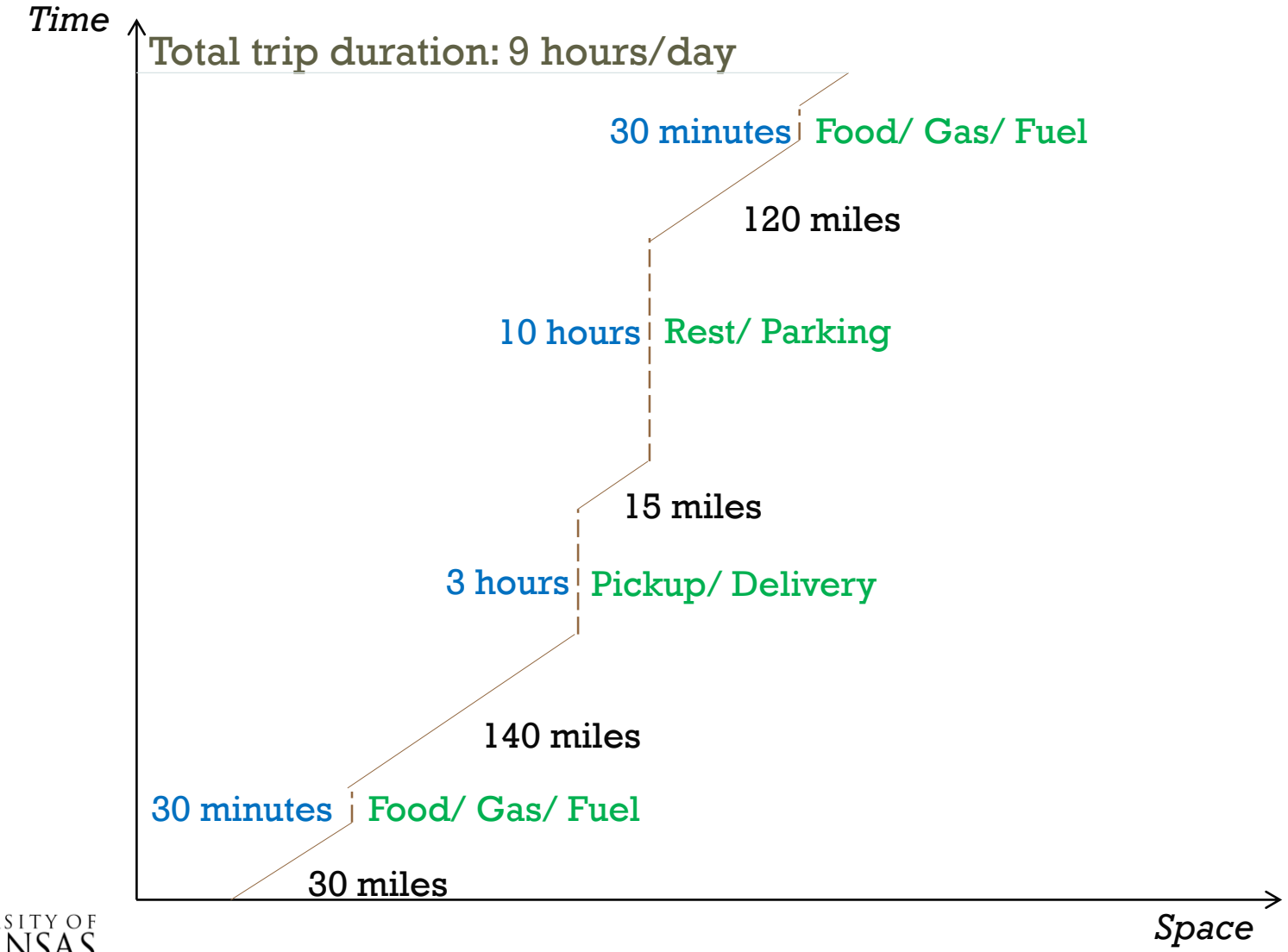
- Shipper characteristics
- Receiver characteristics
- Supply-chain components (e.g., shipment size, commodity type, etc.)
- Business establishments and firms
- Mode choice (i.e., road, rail, water, air, and pipeline)

[2] Data Source

- Business establishment data
- Surveys (e.g., commercial travel survey, vehicle survey, origin destination survey, etc.)
- Commercial fleet GPS data

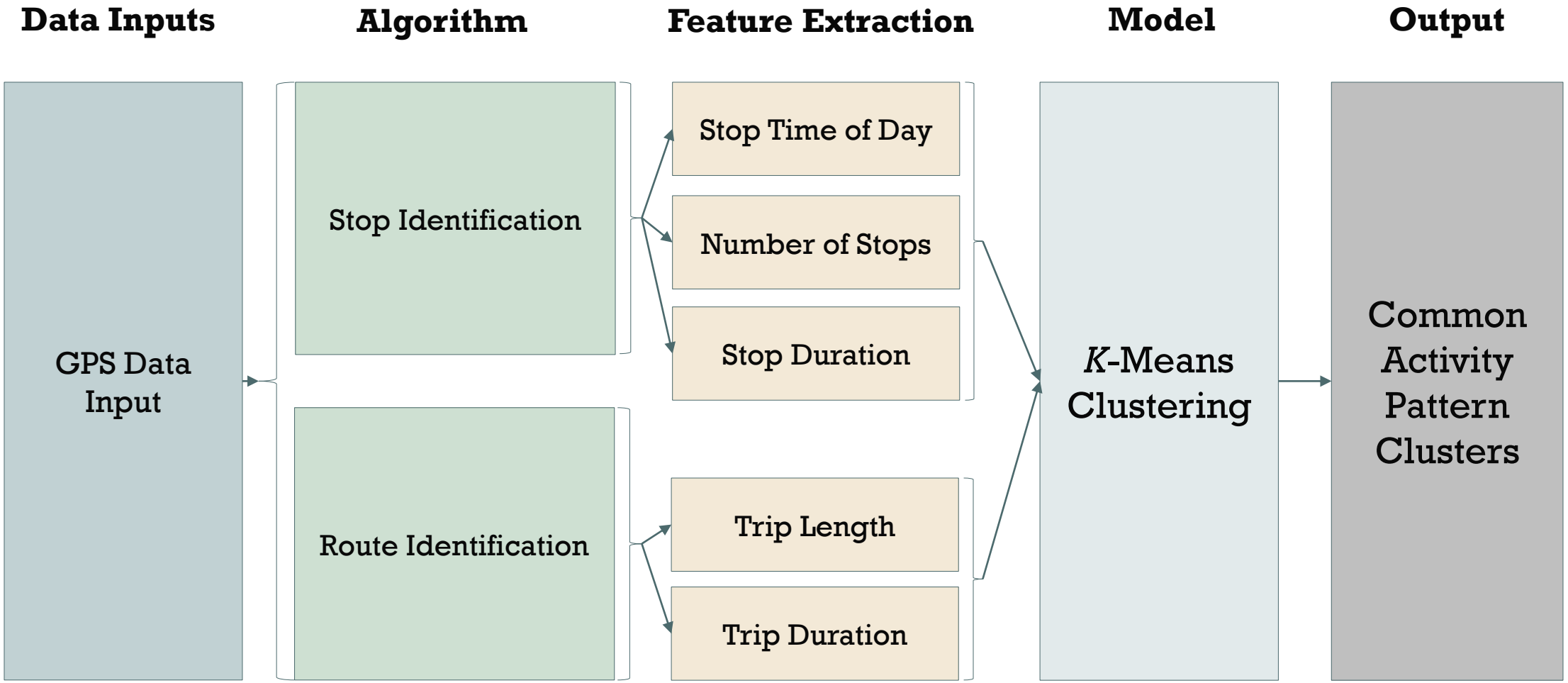
[1, 2]Source: *De Jong and Ben-Akiva, 2007; Roorda, Cavalcante, McCabe, and Kwan, 2010; Jing and Ben-Akiva, 2018*

Activity Pattern Profile Example



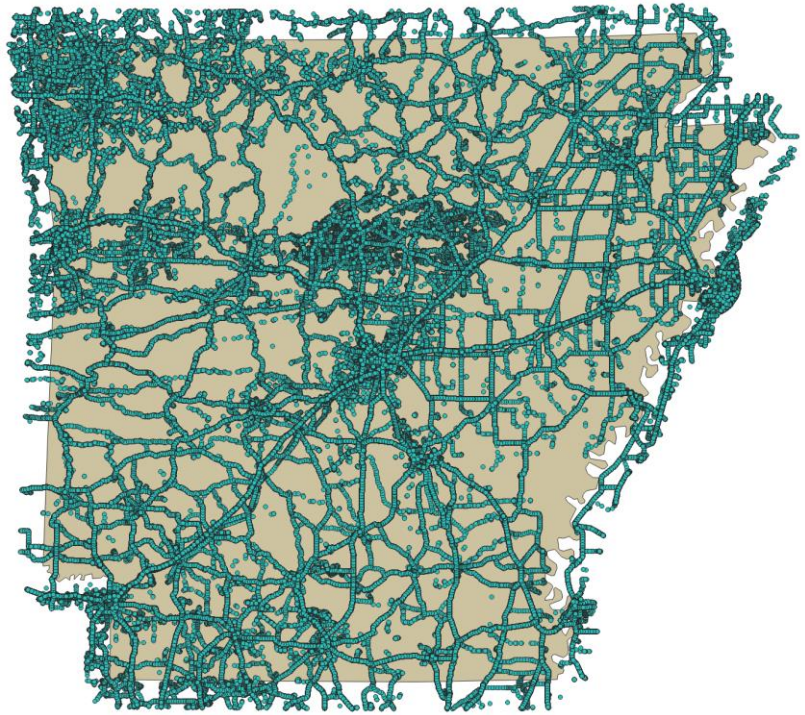
- Tie activity patterns to industry of truck, to use as input for freight travel demand models
- Reduce the number of unique "agents" in ABM to be representative of activity patterns for each industry

Activity Pattern Classification

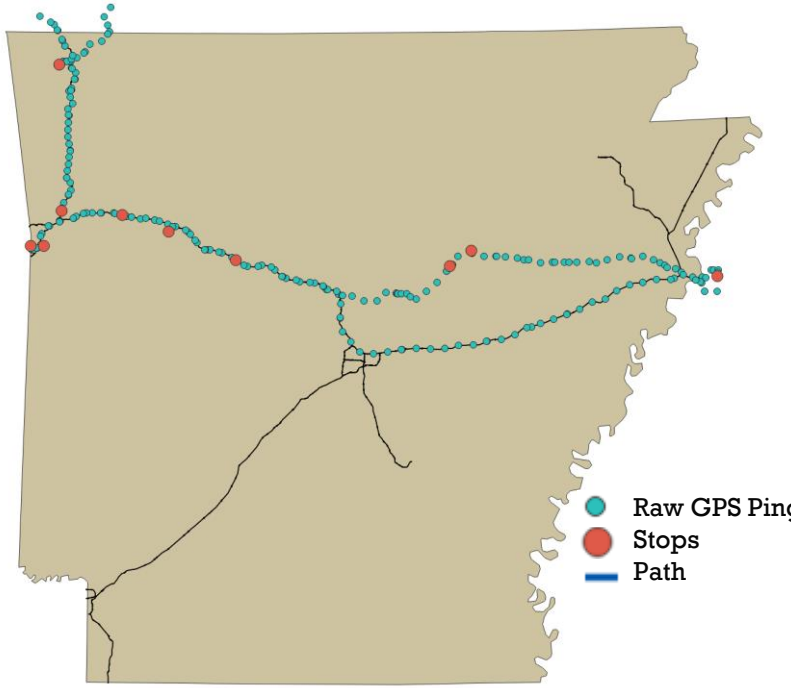


Stops and Paths from a Large GPS Data

Collected GPS Data → Stop Identification → Route Identification



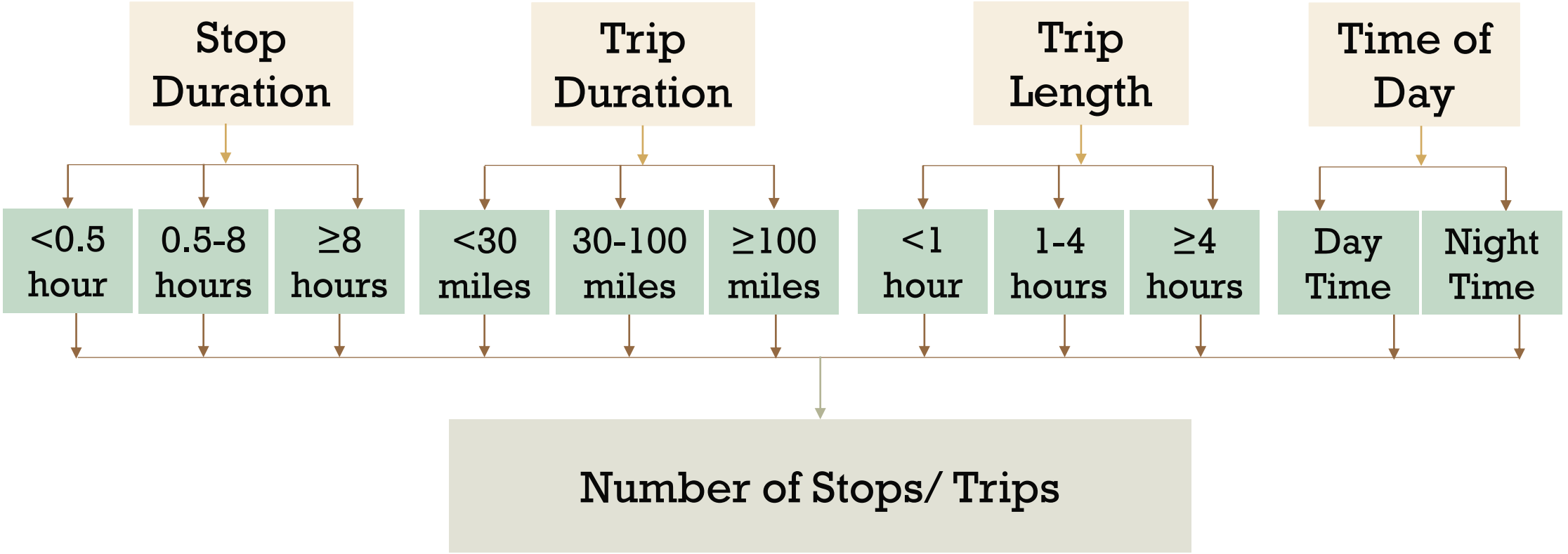
National truck GPS sample
2 week period, 82,000 trucks



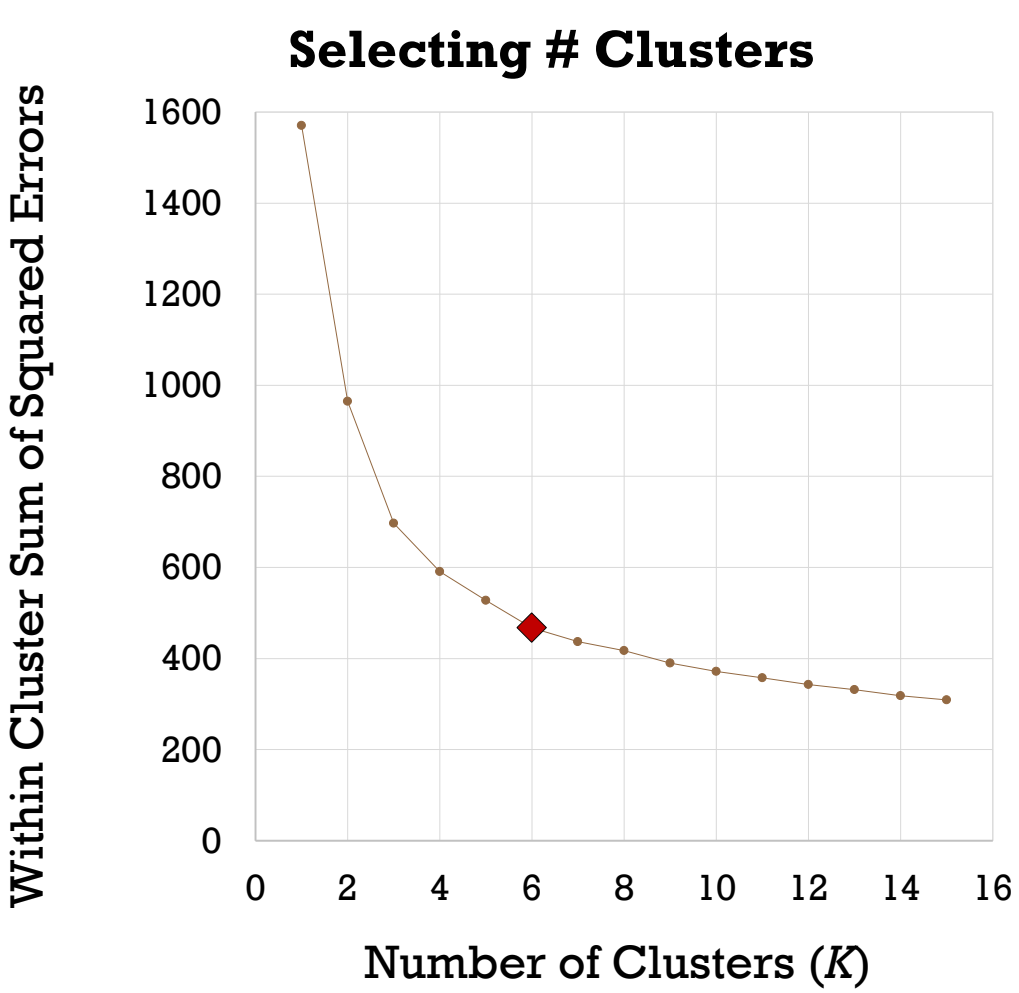
Stops based on speed, duration, and geographical coverage

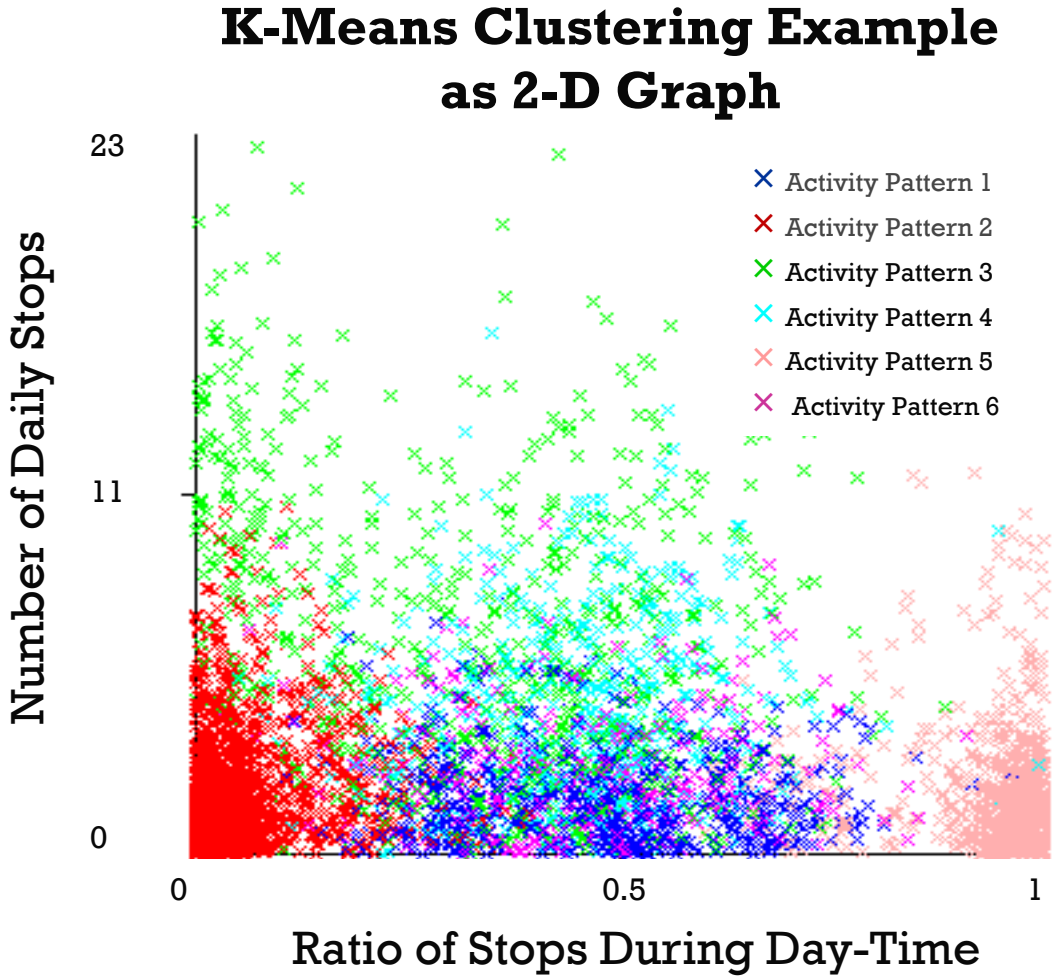
Complete, fully-connected links comprising the truck path

Feature Extraction for Daily Activity Patterns

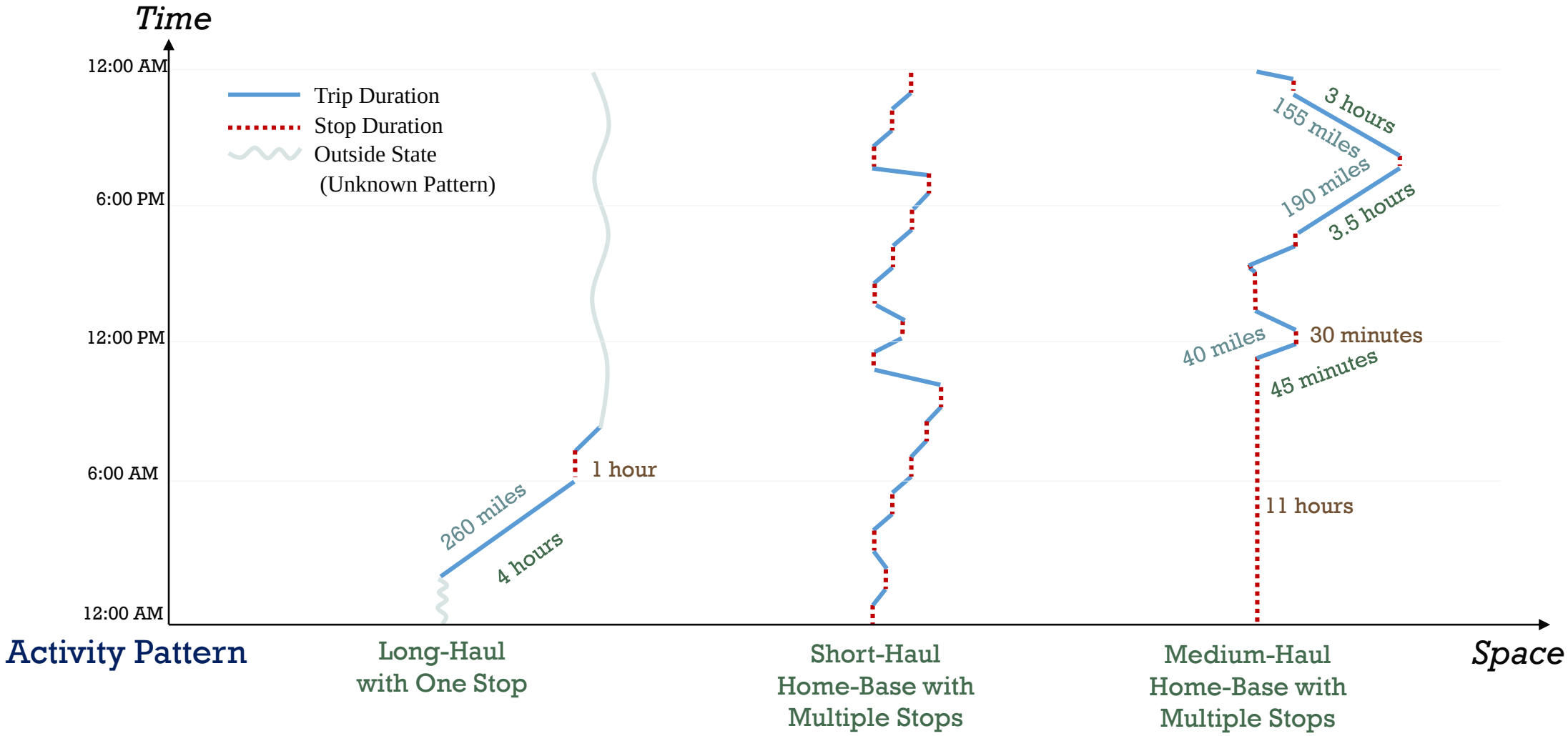


K-Means Clustering

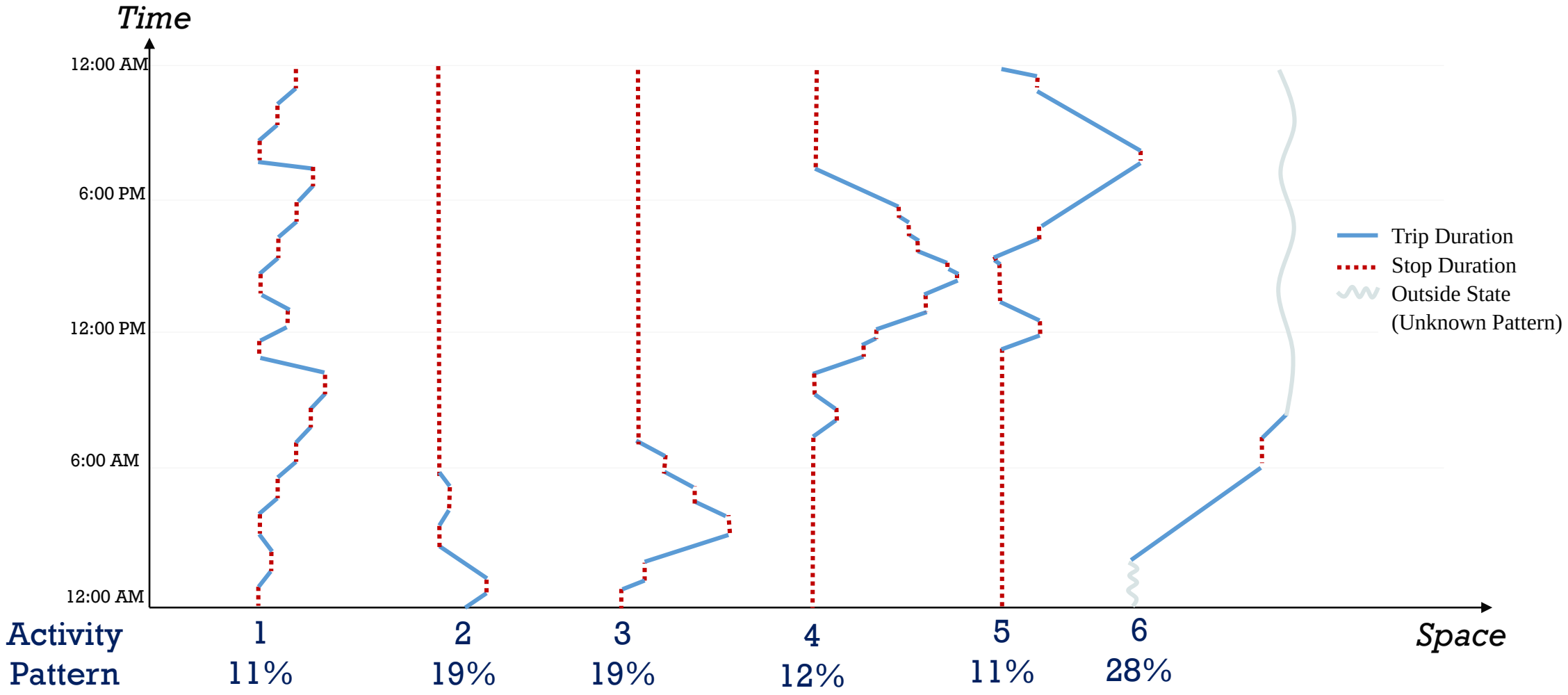




Model Performance: Activity Patterns

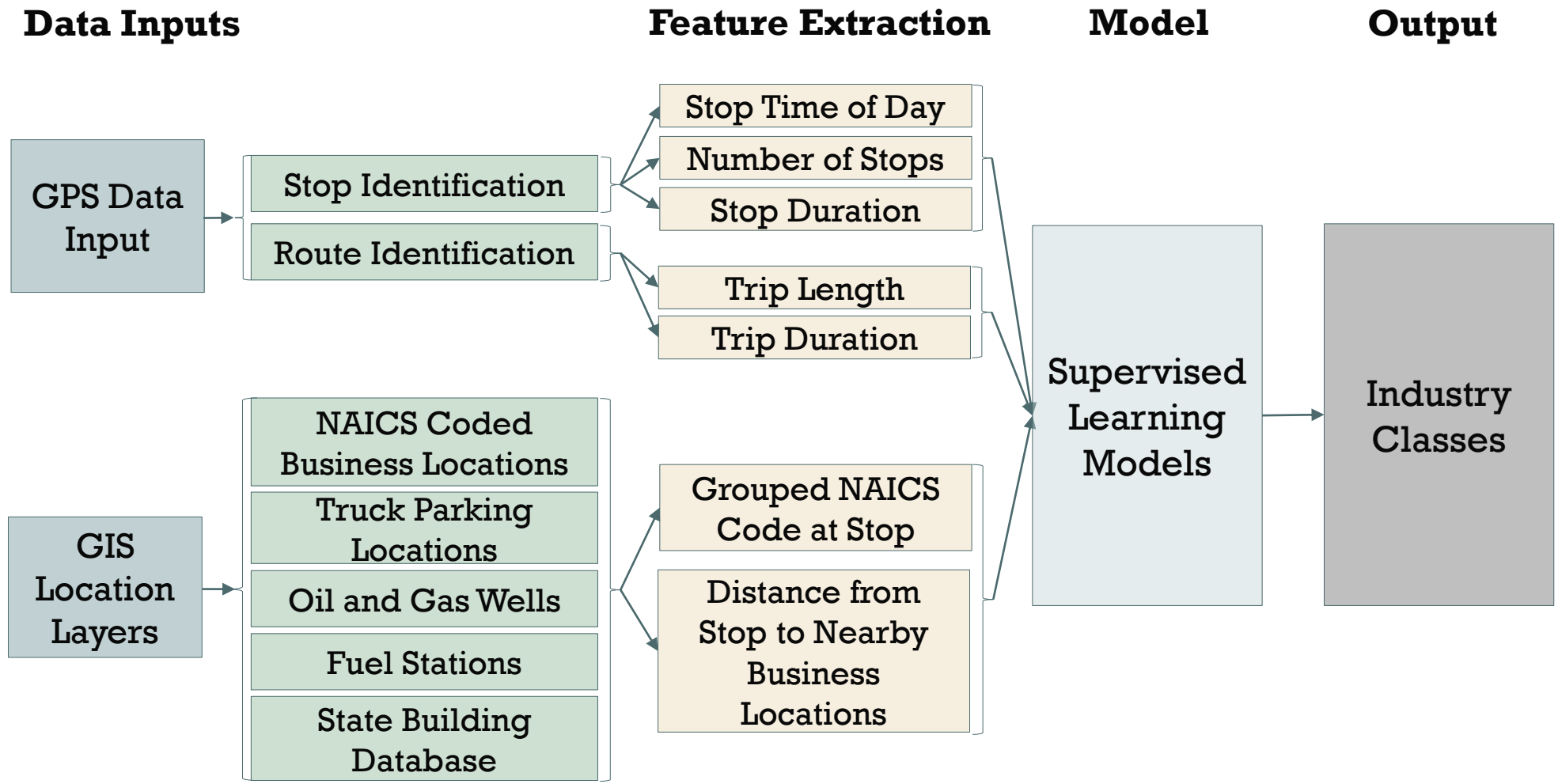


Model Performance: Activity Patterns

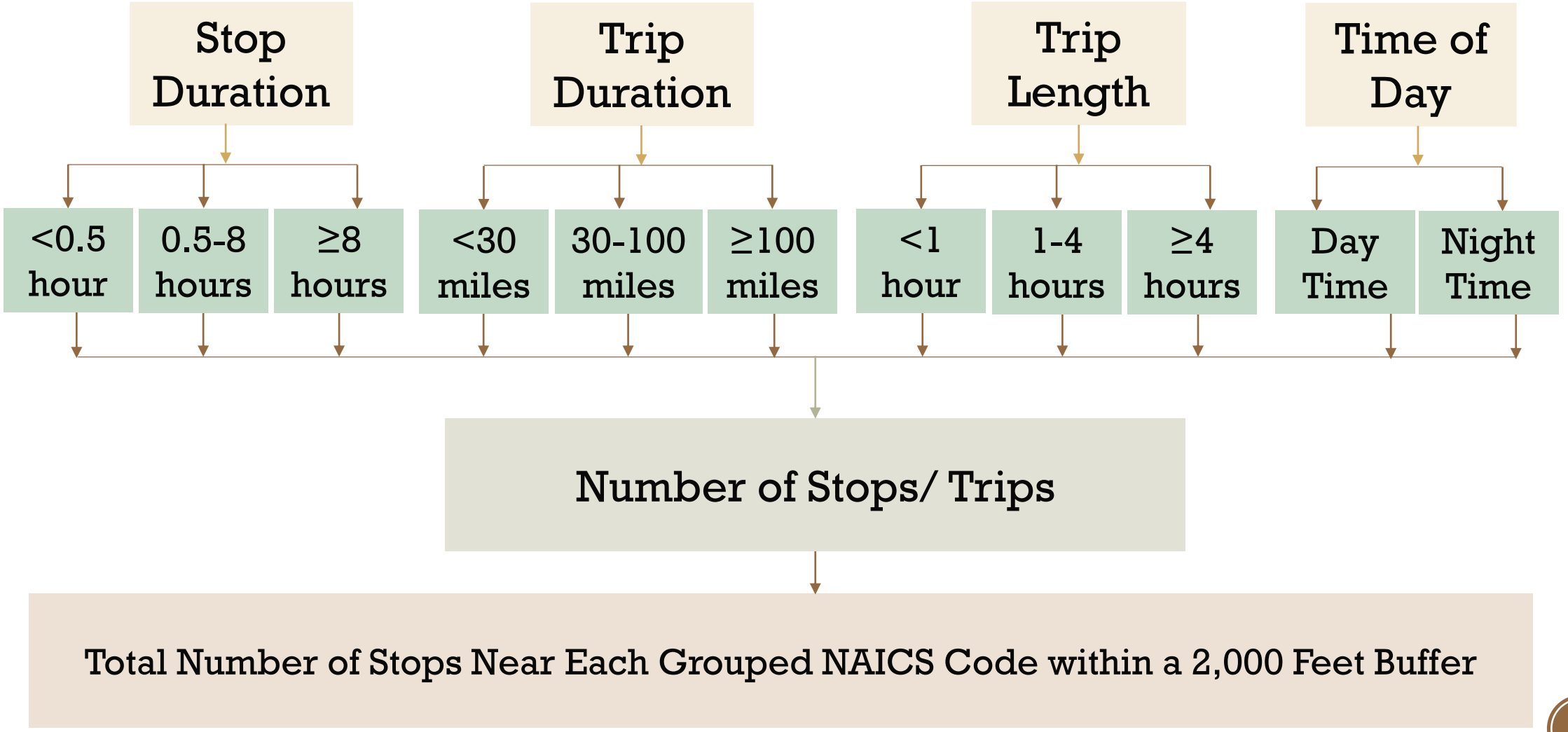


UNIVERSITY OF ARKANSAS

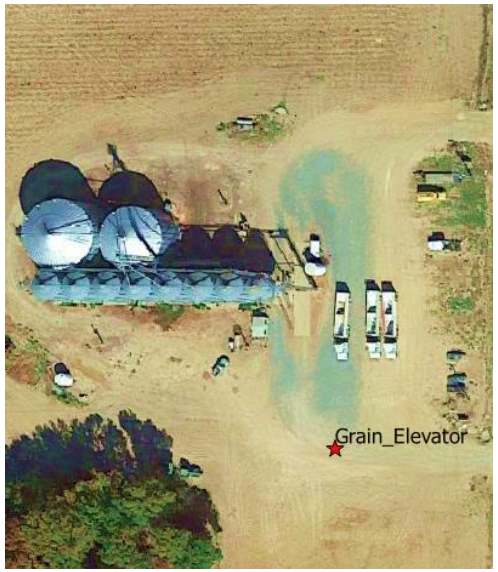
Industry Classification Model



Feature Extraction for Industry Classification



Data Description: Business Layers



Agriculture

Livestock

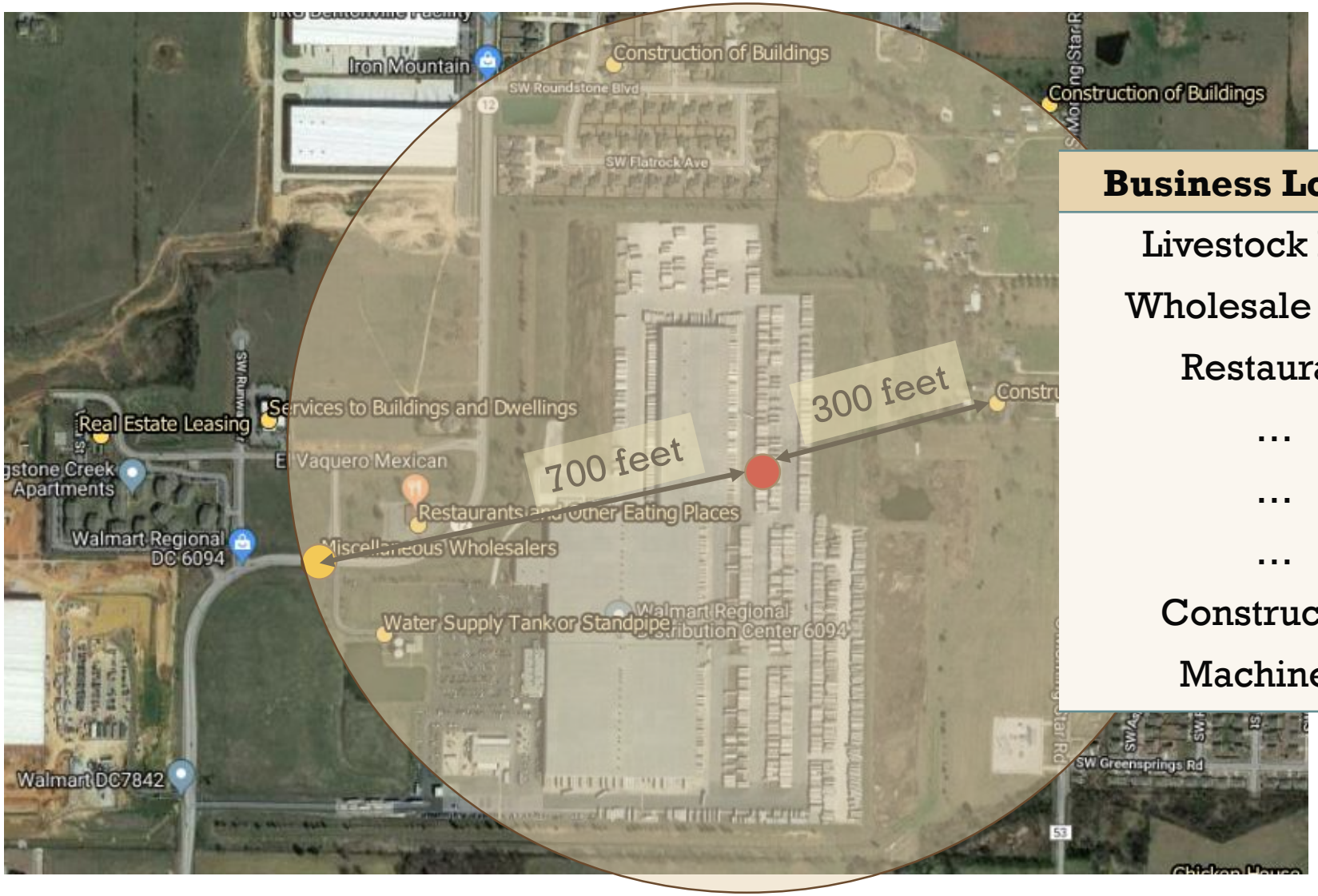
Commercial Food Product

Oil/Gas Well

Department Stores

…and more

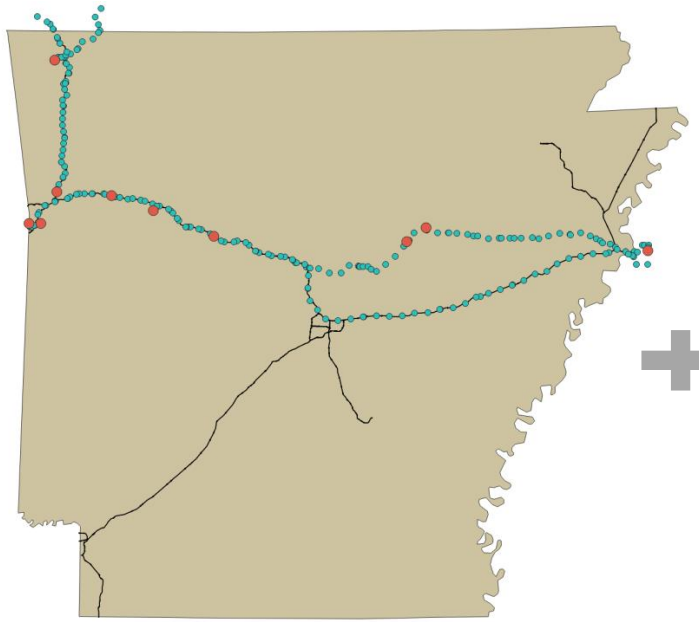
Data Description: Buffer Around Business Layers



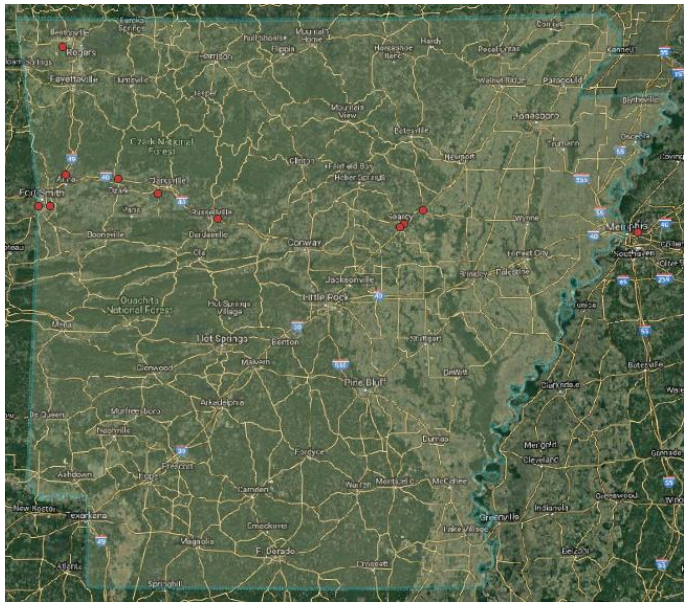
Business Location	Probability
Livestock Farm	0
Wholesale Trade	1
Restaurant	1
…	…
…	…
…	…
Construction	1
Machinery	0

UNIVERSITY OF ARKANSAS

Data Description: Groundtruth Process



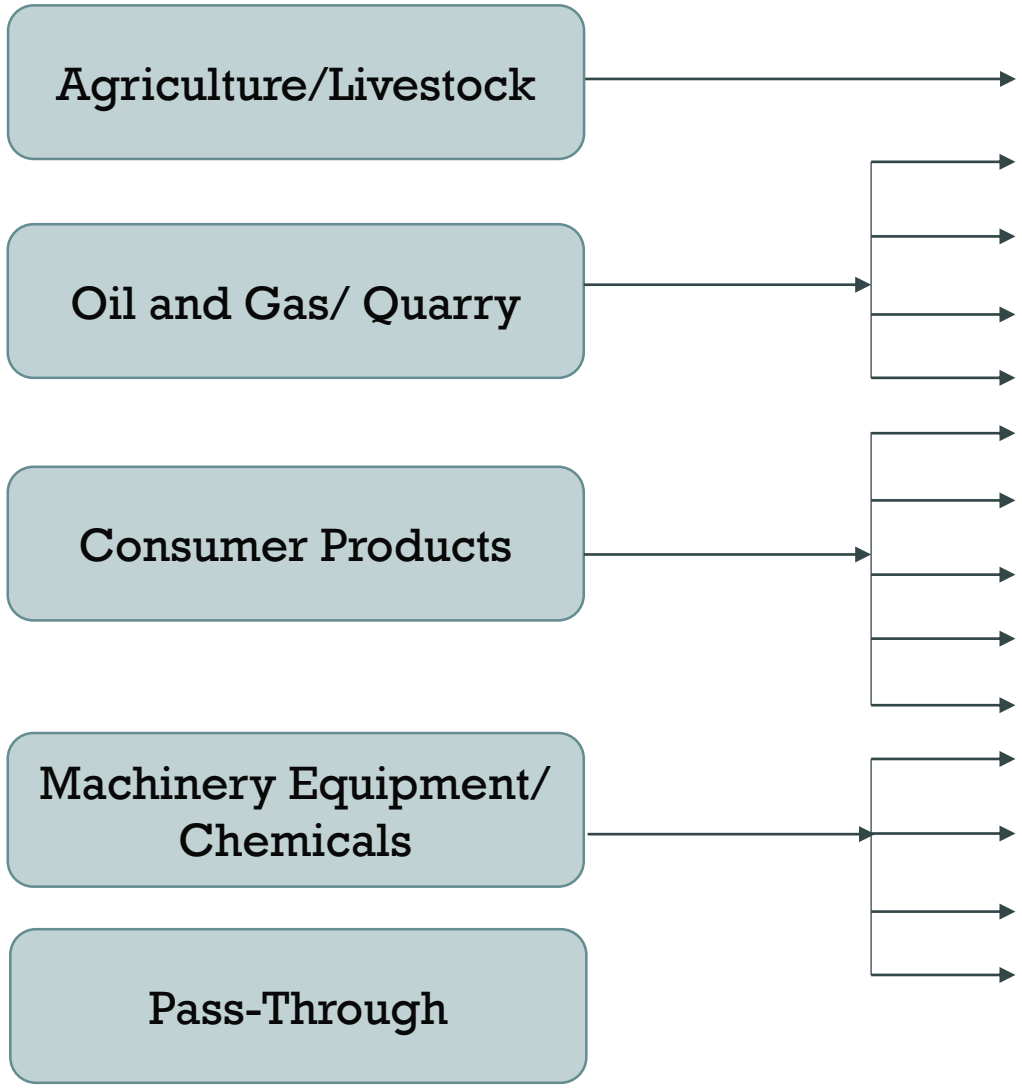
Stop Location



Satellite Image Layer
& Business Layer

Consumer Products

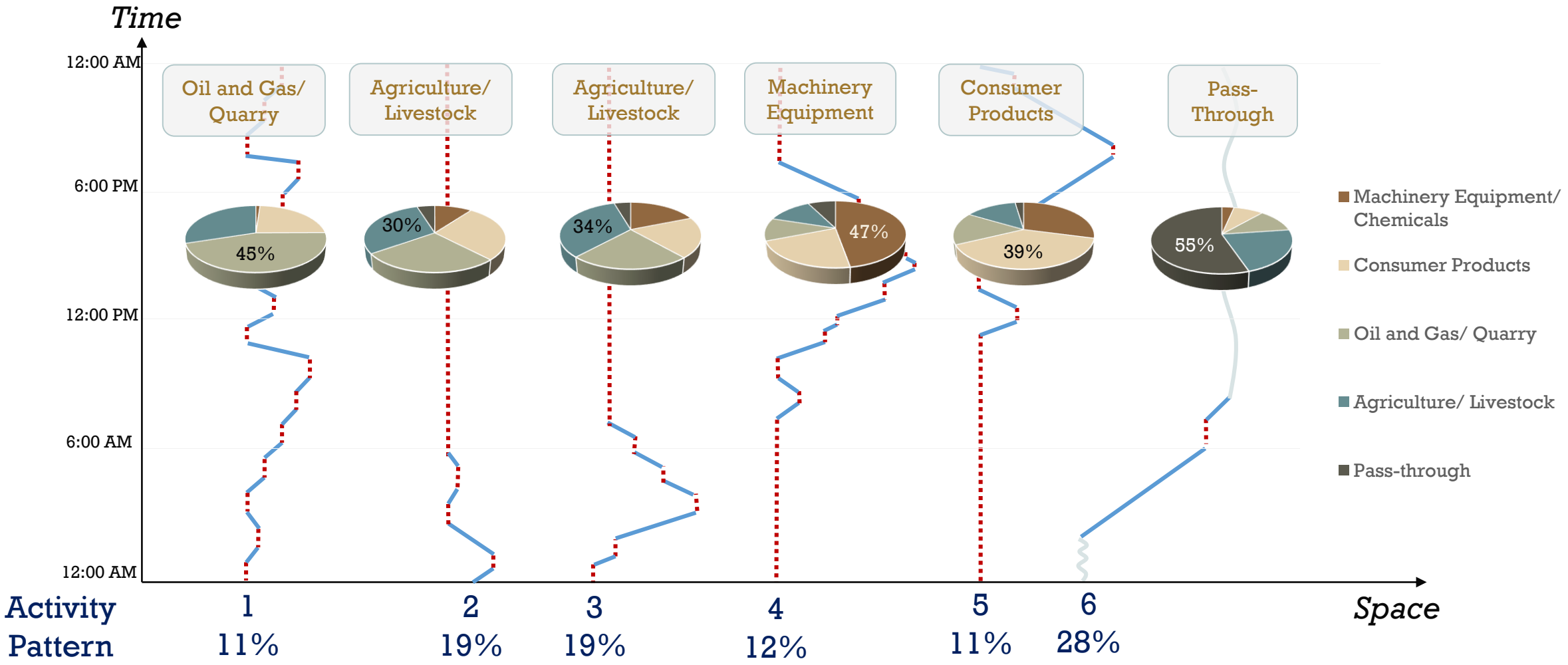
Linking Predicted Industry Class to Commodity



Commodity Name
Agriculture
Mining
Coal
Petroleum
Nonmetallic Minerals
Food
Consumer Manufacturing
Non-Durable Manufacturing
Lumber
Durable Manufacturing
Paper
Chemicals
Clay, Concrete, Glass
Primary Metal
Secondary & Misc. Mixed

Source: Arkansas Statewide Travel Demand Model, 2015

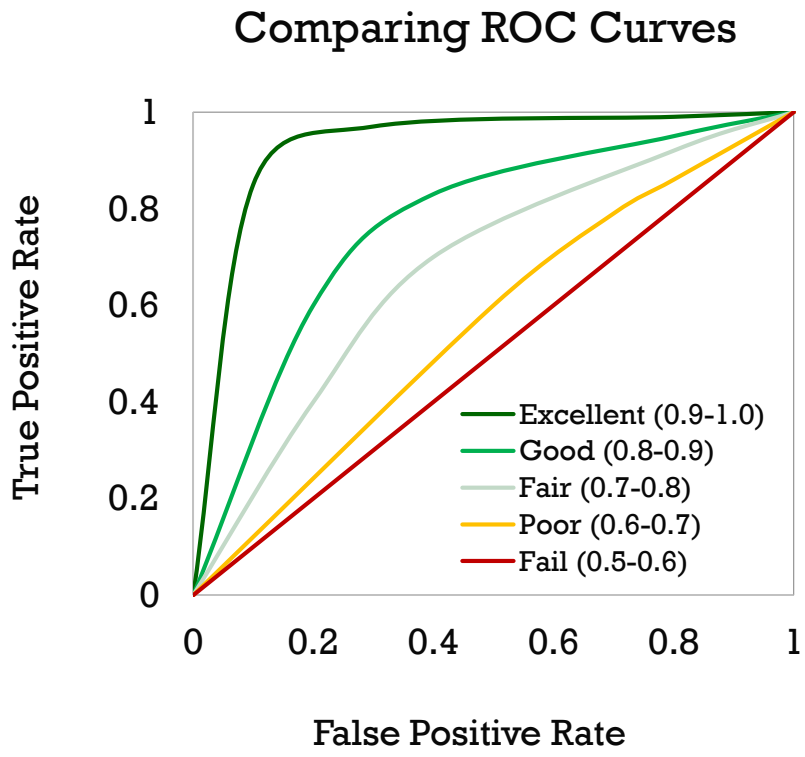
Linking Activity Patterns to Industry



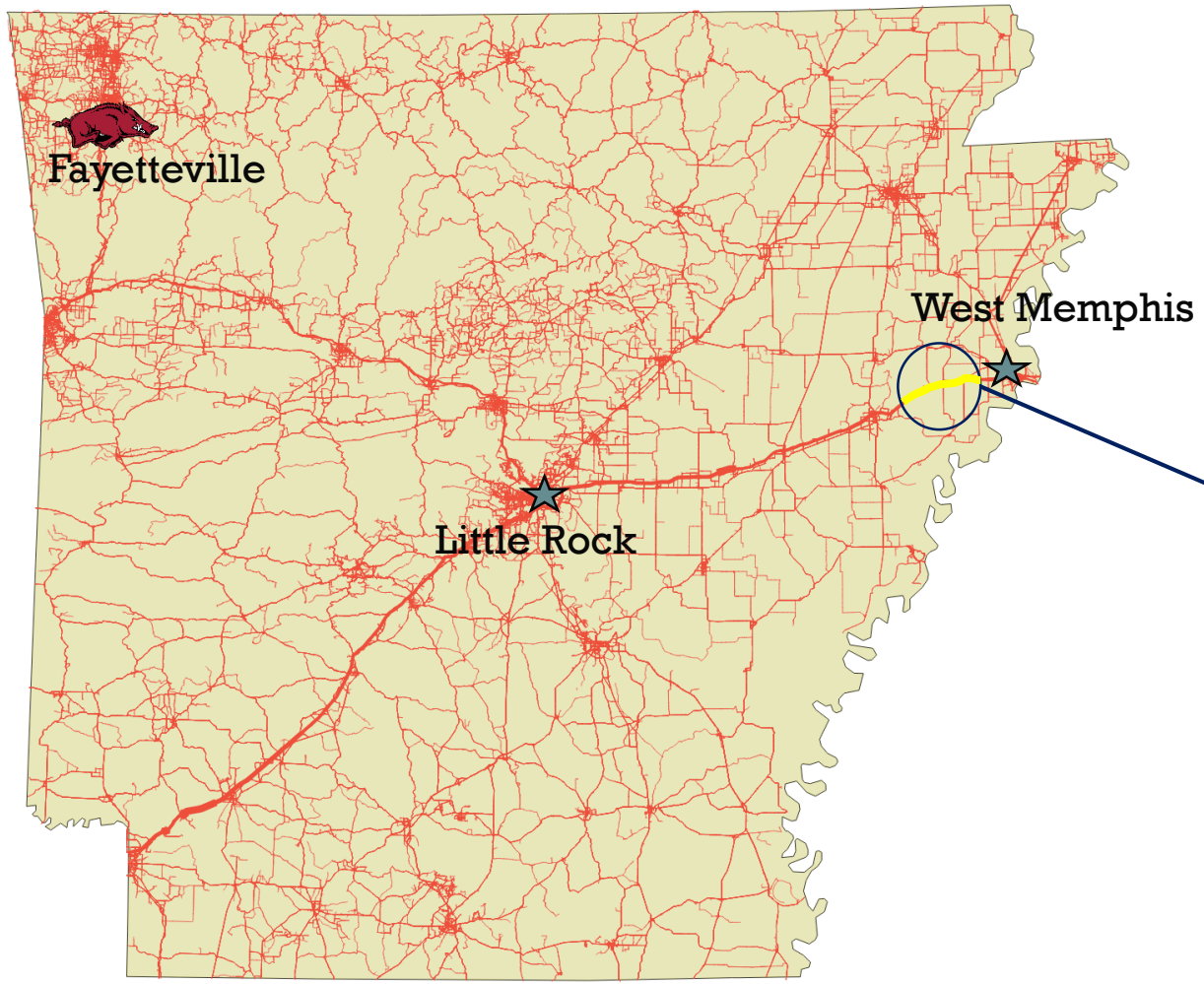
Model Performance: Industry Classification

- Random Forest Machine Learning model implemented
- 3000 samples, ~80/20 training/testing

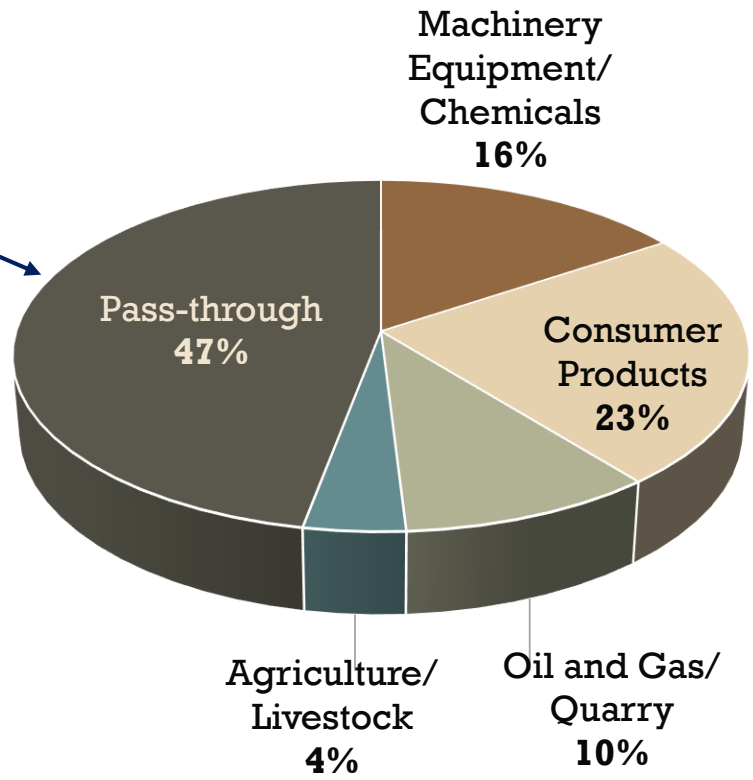
Industry Type	ROC Area	Correctly Classified Instances (%)
Consumer Products	0.91	87
Agriculture/Livestock	0.95	81
Pass-through	0.99	75
Oil and Gas/ Quarry	0.93	73
Machinery Equipment/ Chemicals	0.94	40
Weighted Average	***0.93***	***78***



Industry Classes on Road Links



Percentage of Predicted Industries on A Road Link



Future Work

Different methods for feature extraction

- Principal Component Analysis
- Correlation-Based Method
- Info-Gain Ratio Method
- Wrapper Method

More '*groundtruth*' data to develop the supervised learning model

Application programming interface (API) of Google satellite image to get more accurate location of the activity

Questions?

Thank You

Taslima Akter
PhD Candidate
Department of Civil Engineering
University of Arkansas
Phone: 419-418-1368
Email: takter@uark.edu

Sarah Hernandez, PhD
Assistant Professor
Department of Civil Engineering,
University of Arkansas
Email: sarahvh@uark.edu