UDAT: User Discrimination using Activity-Time Information

Snigdha Das*, Dibya Jyoti Roy*, Subrata Nandi†, Sandip Chakraborty*, Bivas Mitra*

*Dept. of CSE, IIT Kharagpur, India; †Dept. of CSE, NIT Durgapur, India
Email: snigdhadas@sit.iitkgp.ernet.in

CNeRG IIT Kharagpur

Objective

✓ *Developing UDAT, an user identifier cum discriminator model*
- *Leveraging on the daily activity patterns and*
- *Temporal information of the users*
- *In a privacy preserving manner*

Key Idea

✓ Activity pattern followed by an individual carries a signature of that person
✓ **Example:** Alice and Bob's daily routine in the weekdays
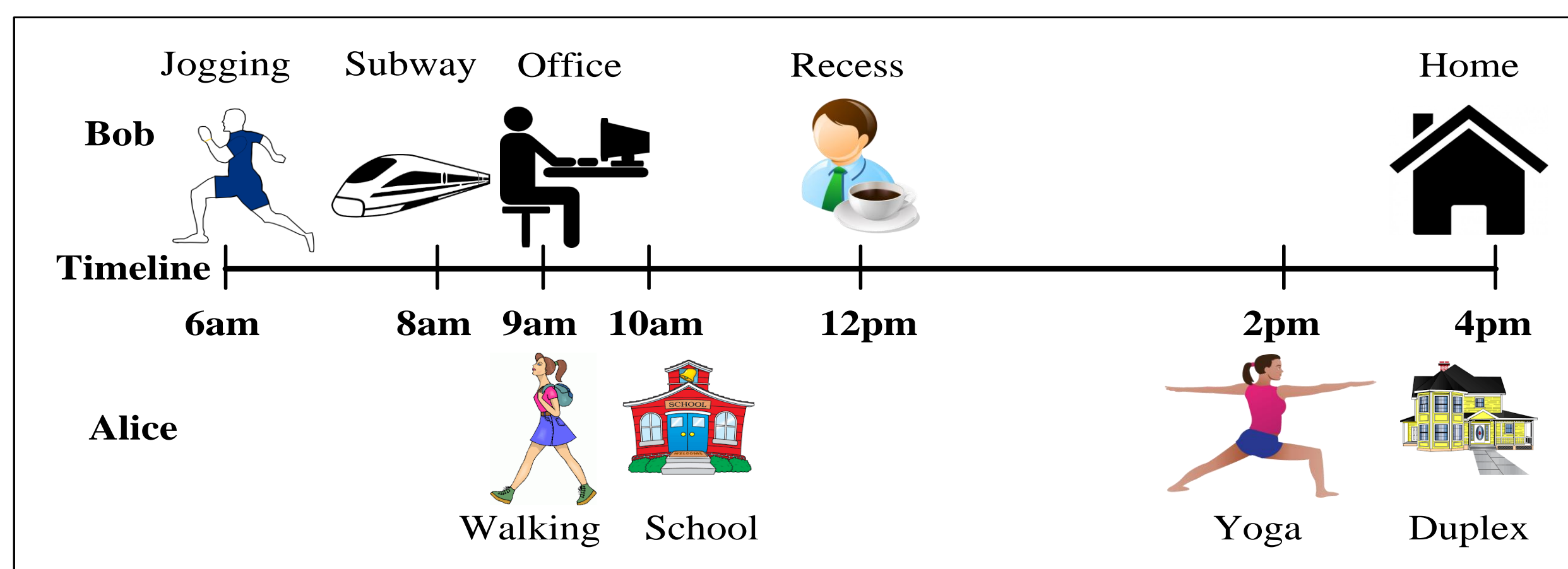
✓ Daily activity patterns of Alice & Bob are **substantially different** in terms of activities and temporal information

Motivation

- **Personalized recommendation** is necessary for the development of the future generation services in the domain of "**Smart City**"
- Smart devices should **automatically recognize** individual users
- **Registering** individual users each time may appear **restrictive**
- System should automatically identify the users in a **privacy preserved manner** without accessing any personally identifiable information (like IMEI) and perform service differentiation

Does daily routine discriminate the users?

✓ Daily routine of three users (Geolife[1] Dataset)
✓ **Major activity of U1:** walking
✓ **Major activity of U2:** biking
- Performed **activities** simply **differentiate** the users U1 and U2
✓ **Biking time of U2:** 0 - 180 (12am - 3am) minutes of the day
✓ **Biking time of U3:** 360 – 540 (6am - 9am) minutes of the day
- Routine of U2 and U3 **varies** significantly **over** the period of **time**
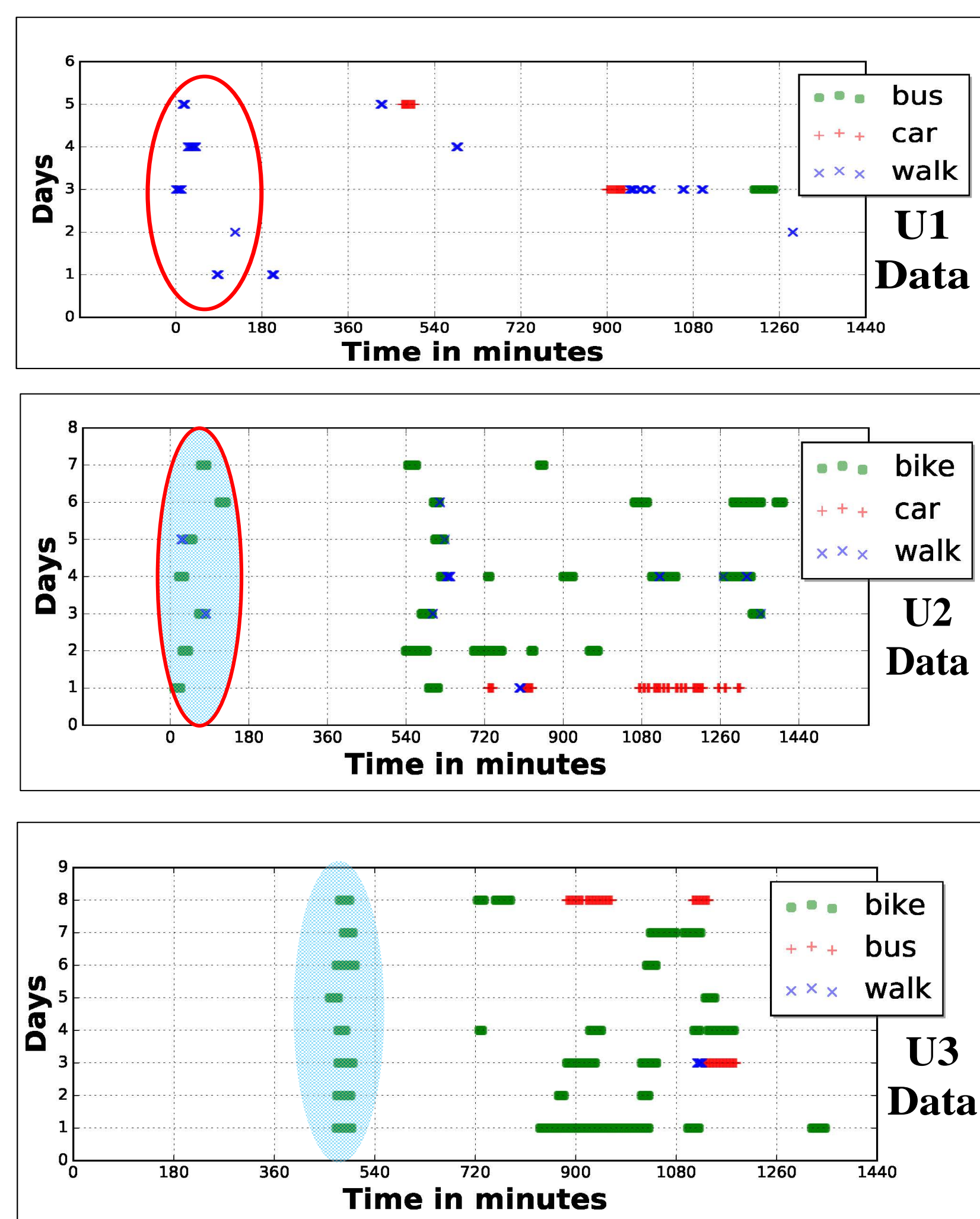
Contributions

✓ **Developing *UDAT*,** a user discrimination model, which involves three major modules – (a) Activity based classification, (b) Outliers detections, and (c) Temporal classification
✓ Developing smartphone based **data collection framework**, named as **UDAT Dataset**
✓ *UDAT* model discriminates users with **73.3%** and **80.68% accuracy,** for Microsoft Geolife and UDAT datasets respectively

Dataset

✓ **Geolife[1] Dataset:**
- Users: 24 out of 178 labelled
- Time: 4 years
- Transportation mode: Bike, Bus, Car, Subway, and Walk
✓ **UDAT Dataset:** Launch experiment at IIT Kharagpur
- Users: 15
- Time: 3 months
- Activity: InVehicle, OnFoot, OnBicycle, Running, Walking, Still, Tilting, and Unknown

UDAT Model

# of Activity Bucket: N
# of User in each Bucket i: i_m
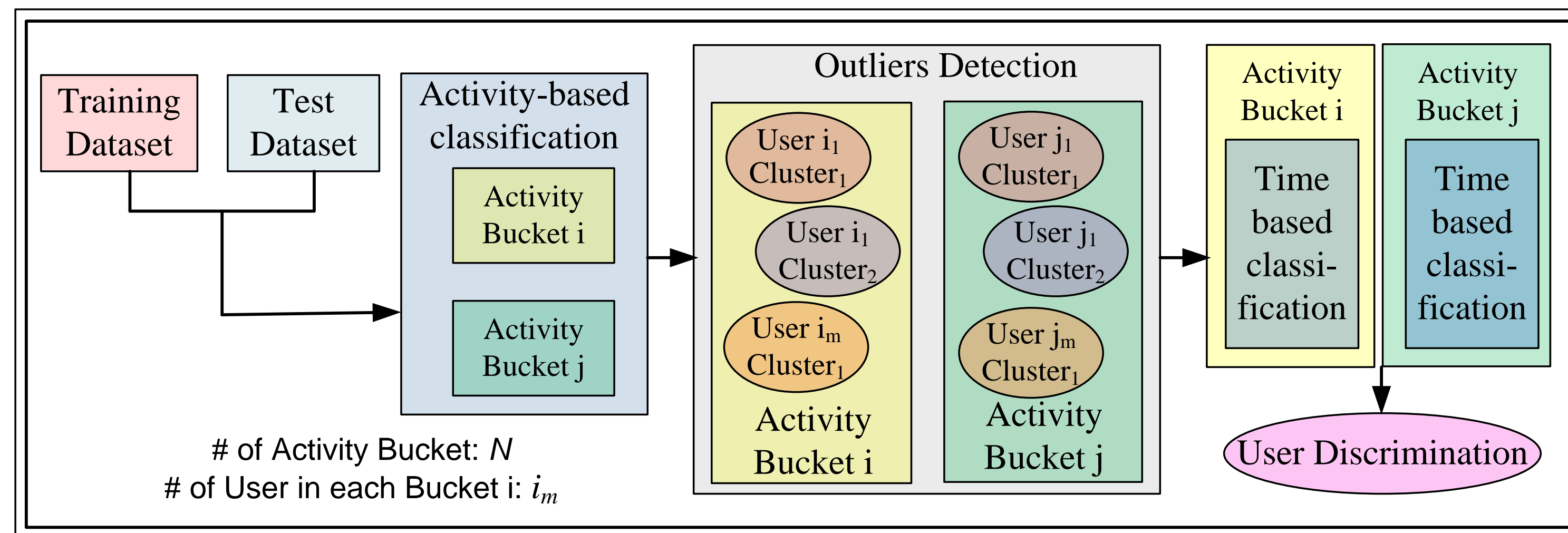
✓ **Activity based Classification:**
- The module classifies users based on the activities performed by them
- It construct one bucket for each activity
- Each activity bucket contains information of several users
✓ **Outliers Detection:**
- Activity bucket holds normal activity pattern along with few exception
- This module eliminates those exception using **DBSCAN** model
- It identifies the users' multiple patterns of activity
✓ **Temporal Classification:**
- The module discriminates the users inside each activity bucket
- **Models:** kNN, Random Forest, Logistic Regression, SVC
- **Temporal features:**
  i. First start time of the activity
  ii. Total duration for that activity
  iii. Maximum activity duration
  iv. Activity trip count

Experimental Results

GEOLIFE DATASET: ACCURACY (%) COMPARISON OF DIFFERENT MODELS (**KM:** Kwapisz's MODEL)					
Activity	*UDAT*	**KM**	**Activity**	*UDAT*	**KM**
Bike	85.47	69.47	Bus	82.92	65.03
Car	83.20	79.46	Subway	70.50	68.68
Walk	44.41	32.44	**Average**	**73.30**	63.01

UDAT DATASET: ACCURACY (%) COMPARISON OF DIFFERENT MODELS			
Activity	*UDAT*	**Kwapisz's Model**	*UDAT_Outlier*
Tilting	NA	39.29	31.13
InVehicle	NA	32.65	33.82
OnBicycle	100.00	22.36	48.03
OnFoot	65.07	32.92	40.78
Running	NA	32.77	NA
Still	84.64	33.16	46.75
Unknown	73.01	24.56	41.61
Walking	NA	56.86	NA
Average	**80.68**	34.32	40.35

Accuracy chart of UDAT dataset using different classifier

✓ **Geolife[1] DataSet**
- *UDAT* Model accuracy – **73.3%**
- *Kwapisz's*[2] Model accuracy – 63%

✓ **UDAT DataSet**
- *UDAT* Model accuracy – **80.68%**
- *Kwapisz's*[2] Model accuracy – 34.32%
- *UDAT_Outlier* Model accuracy – 40.35%
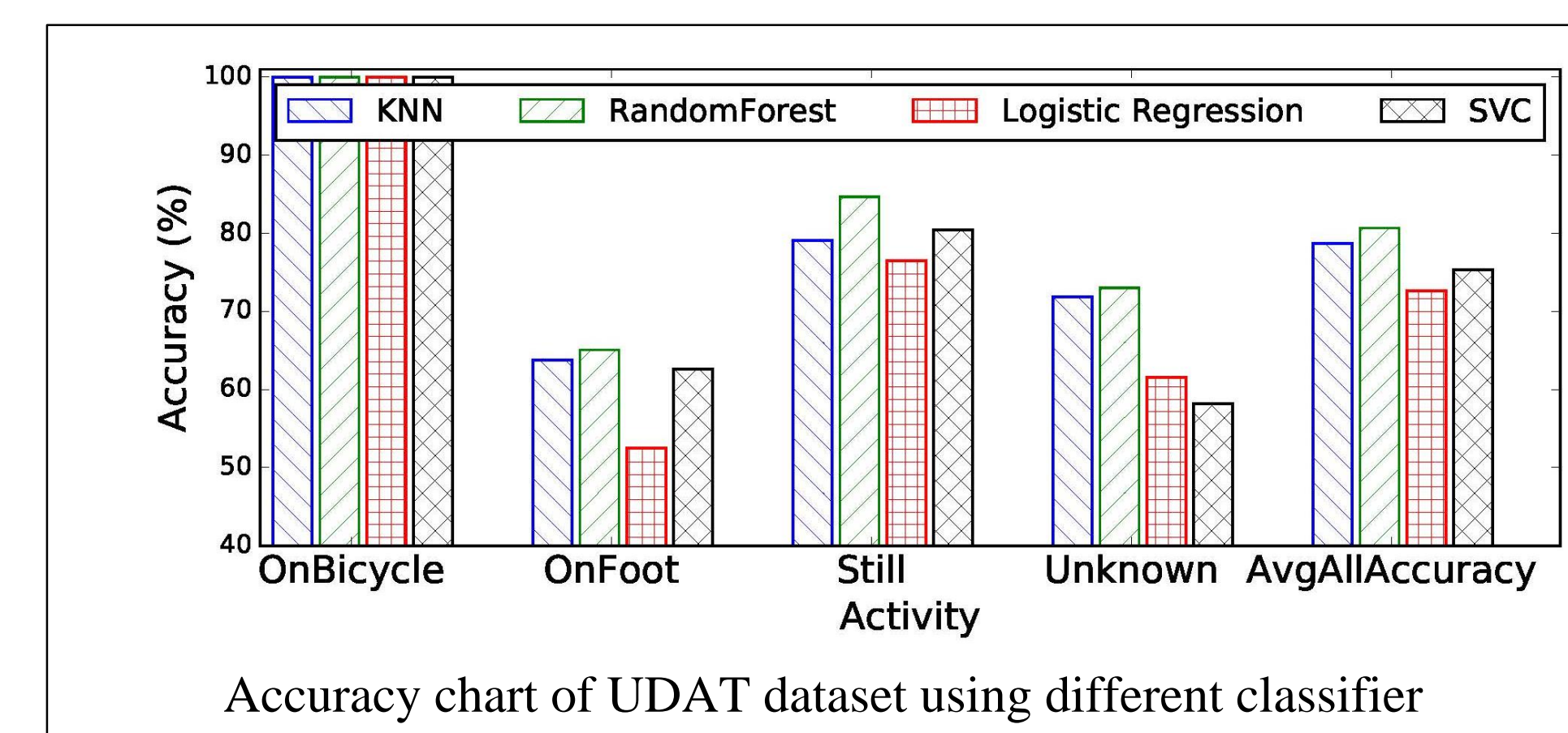
✓ **Different ML Algorithms**
- Average accuracy for all activities > **70%**
- **Random Forest** surpasses others

Conclusion

✓ Demonstrate **activity signatures** as a **valid alternative** to sensor-driven user identification paradigm
✓ Developed *UDAT* model **discriminates** the users using the daily activity patterns
✓ Experimental results state that *UDAT* model **outperformed** the sensor-driven competing algorithms for both Microsoft *Geolife*[1] and UDAT dataset

References

1. Y. Zheng, L. Wang, R. Zhang, X. Xie, and W.-Y. Ma, "Geolife: Managing and understanding your past life over maps," IEEE Mobile Data Management, 2008, pp. 211–212.
2. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification," IEEE Biometrics: Theory Applications & Systems, 2010, pp. 1–7.