Fake AD Click Detection by Identifying Anomalies

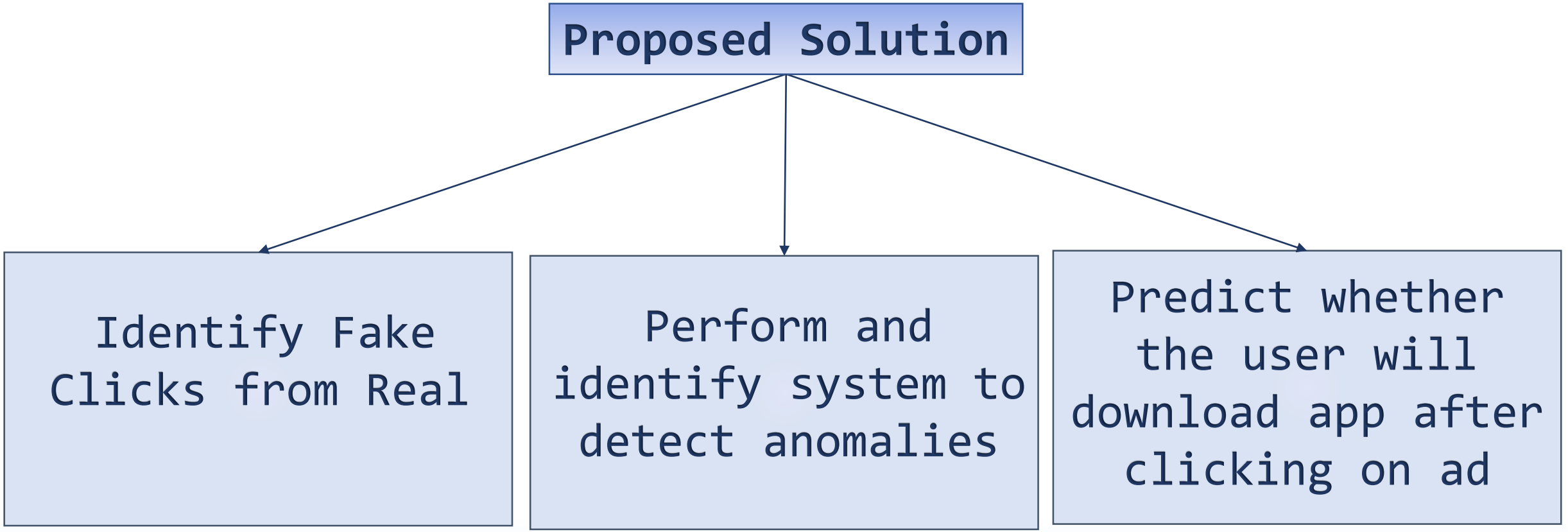
From: Gautam Shanbhag

(18210455)

BACKGROUND

Click fraud is a black-hat technique of falsely inflating the number of clicks on a pay-per-click ad.

- ✓Advertisers are trying to sabotage their competitors by driving up their costs and meeting their budget caps early on in the day
- ✓Ad publishers are clicking on the ads displayed on their own sites to generate more revenue for themselves.


Proposed Solution
Identify Fake Clicks from Real
Perform and identify system to detect anomalies
Predict whether the user will download app after clicking on ad

LITERATURE REVIEW

✓*"Prediction of click frauds in mobile advertising – IC3 2015"*

- *feature selection using Recursive Feature Elimination (RFE)*
- *classification through Hellinger Distance Decision Tree (HDDT)*
- $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$
- *accuracy achieved by proposed framework is 64.07 %*

✓*"Exposing click-fraud using a burst detection algorithm – IEEE 2011"*

- *Set of webpages selected and random visited*
- *No. of visits calculated between user visits*
- *Splay tree to store webpages/ip*
- *Realtime usage*

LITERATURE REVIEW

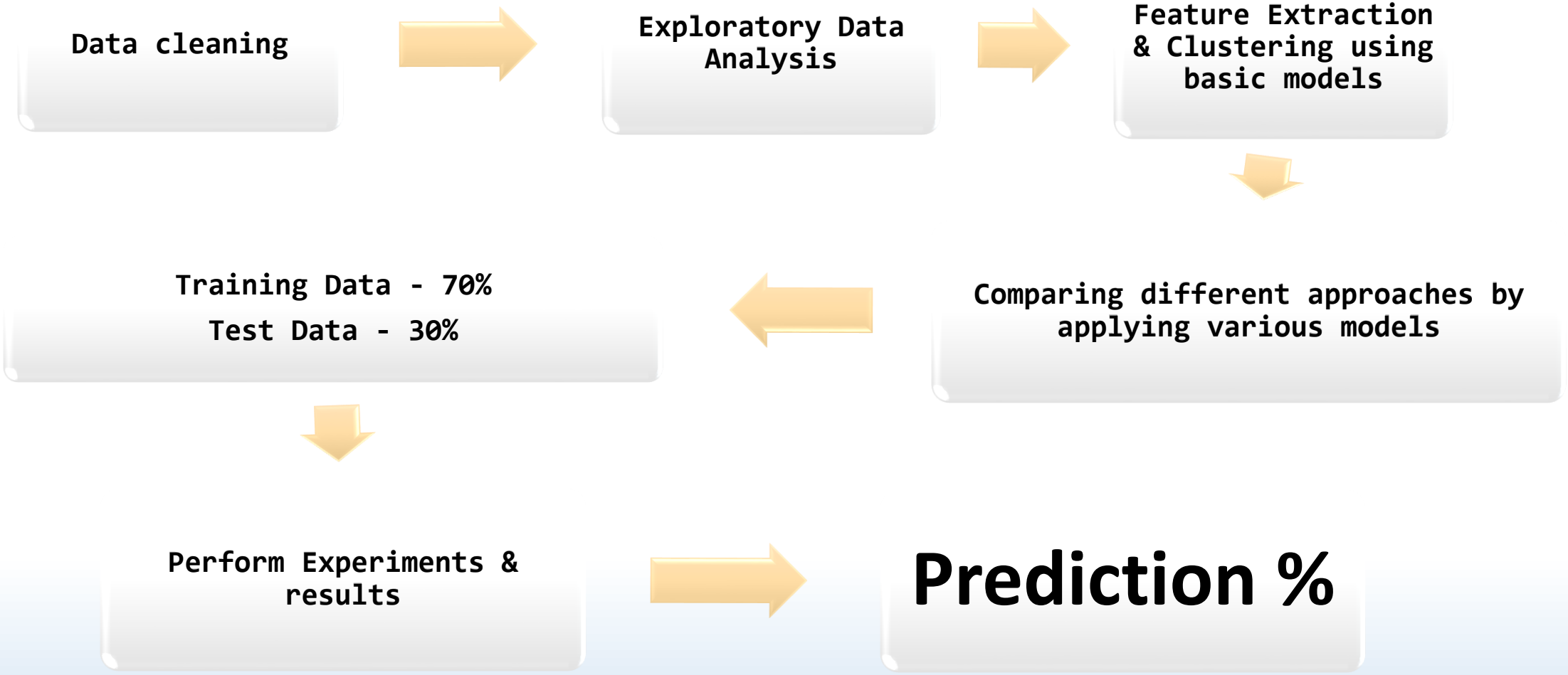
- ✓*"Real Time Click Fraud Prevention using multi-level Data Fusion – WCECS 2010"*
  - *Dempster-Shafer evidence theory*
  - *multi level data fusion mechanism*
- ✓*"Detecting insurance claims fraud using machine learning techniques – ICCPT 2017"*
- ✓*"Machine learning algorithms for document clustering and fraud detection – ICDSE 2016"*

DATASET

- ✓Dataset provided by TalkingData – China's Mobile Big data platform
- ✓More than 1 million records
- ✓Attributes like

  id, timestamp, site_id, site_domain, app_id, app_category, device_id, device_ip, device_model, device_type, and other10 categorical values, **is_attributed**.

PLAN


Data cleaning
Exploratory Data Analysis
Feature Extraction & Clustering using basic models
Comparing different approaches by applying various models
Training Data - 70%
Test Data - 30%
Perform Experiments & results
Prediction %

Questions ?

References

- [1] M. Taneja, K. Garg, A. Purwar and S. Sharma, "Prediction of click frauds in mobile advertising," 2015 Eighth International Conference on Contemporary Computing (IC3), Noida, 2015, pp. 162-166. doi: 10.1109/IC3.2015.7346672
- [2] D. Antoniou et al., "Exposing click-fraud using a burst detection algorithm," 2011 IEEE Symposium on Computers and Communications (ISCC), Kerkyra, 2011, pp. 1111-1116. doi: 10.1109/ISCC.2011.5983854
- [3] C. Walgampaya, M. Kantardzic, R. Yampolskiy, "Real Time Click Fraud Prevention using multi-level Data Fusion", WCECS 2010, October 20-22, 2010, San Francisco, USA
- [4] R. Roy and K. T. George, "Detecting insurance claims fraud using machine learning techniques," 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), Kollam, 2017, pp. 1-6. doi: 10.1109/ICCPCT.2017.8074258
- [5] S. Yaram, "Machine learning algorithms for document clustering and fraud detection," 2016 International Conference on Data Science and Engineering (ICDSE), Cochin, 2016, pp. 1-6. doi: 10.1109/ICDSE.2016.7823950
- [6] K C Wilbur, Y Zhu, D S. Anderson, "Click Fraud[J]", Access & Download Statistics, vol. 28, no. 2, pp. 293-308, 2009.
- [7] Urbanski Al., (01 May 2013). “Bots Mobilize”, DMN [Online]. Available: http://www.dmnews.com/bots-mobilize/article/291566/.