related issues are found in [7] which needs to provide correct recommendations.

Many recommendation models are studied in [8] using content based collabrotie filtering and other techniques. These techniques are meant for providng genuine recommendations for different mobile applications. A framework was proposed in [9] for context-aware recommendations which makes use of dependency assumptions and context independency. In this paper we proposed a framework that makes use of mobile app work flow and historical knowledge as inputs and perform ranking fraud detection. The detection technique uses leading sessions and perform collection of ranking based, review based and rating based evidences and aggregate them in order to identify fraud.

The remainder of the paper is structured as follows. Section II provides review of literature. Section III presents the proposed system in detail. Section IV presents experimental results while section V concludes the paper.

## 2. PROPOSED SYSTEM

This section presents the proposed framework that guides in discovering ranking fraud of mobile applications in the real world. The proposed framework takes leading sessions of mobile apps as input and perform discovery of fraud evidences. The evidences are of three types. They are ranking based evidences, rating based evidences and review based evidences. These evidences are then aggregated in order to have business intelligence.

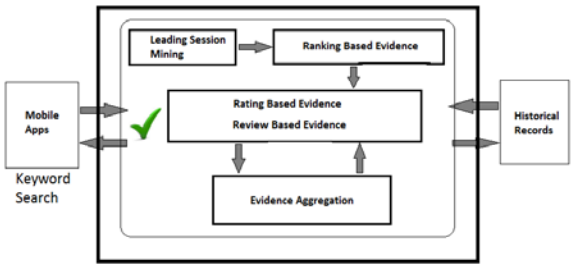


Figure 1: Proposed Framework

In the process, the framework makes use of historical results that gives needed knowhow to identify fraud related evidences. The evidence aggregation process and other phases of the proposed system are bestowed with heuristics to make well informed decisions. The proposed system also makes use of local and global anomaly approaches in order to determine the fraud instances.

## 3.IMPLEMENTATION AND EXPERIMENTAL RESULTS

We built a prototype application to demonstrate the proof of concept. The application is built using Java platform. The application is web based and takes inputs as discussed in the previous section and detects fraudulent rankings. The functionalities of the system are categorised into three modules namely local anomaly, user and global anomaly.

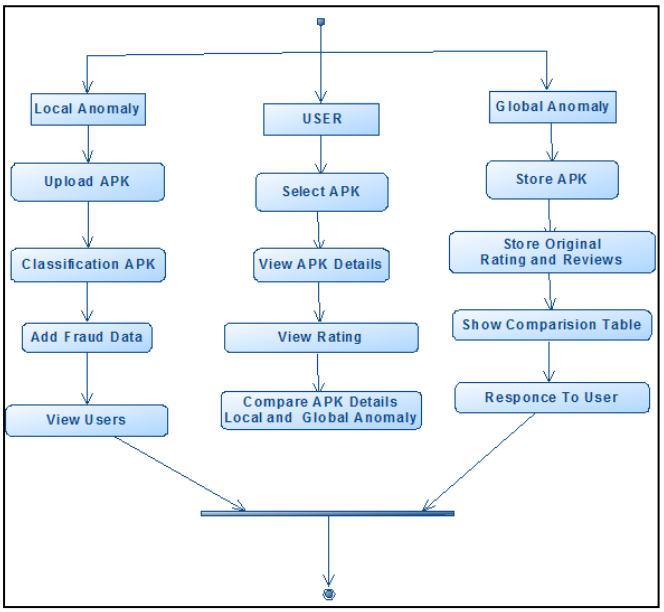


Figure 2: Flow Chart Showing System Functionalities

As shown in Figure 2, the local anomaly module is responsible for uploading APK file, perform classification and find out fraud data. This module is also responsible to view users. The user module is responsible for selection of APK, view APK details, view rating detils and compare APK details for local and global anomaly in the context of the proposed framework. The global anomaly module is responsible to store APK, store original rating and reviews, show comparison tables, and show response to user.

| **Percentage of Full data(%)** | Principal1 | Principal2 |
|---|---|---|
| 10 | 130 | 100 |
| 20 | 40 | 35 |
| 30 | 25 | 20 |
| 40 | 15 | 14 |
| 50 | 8 | 7 |
| 60 | 6 | 6 |
| 70 | 4 | 4 |
| 80 | 2 | 2 |
| 90 | 0 | 0 |
| 100 | 0 | 0 |

**Table 1:** Shows Percentage of Full Data vs. Principal 1 and Principal 2

As shown in Table 1, the results revealed the data percentage with respect to principal 1 and principal 2. The data percentage shows values from 10 to 100 while the principals show values in decreasing trend.

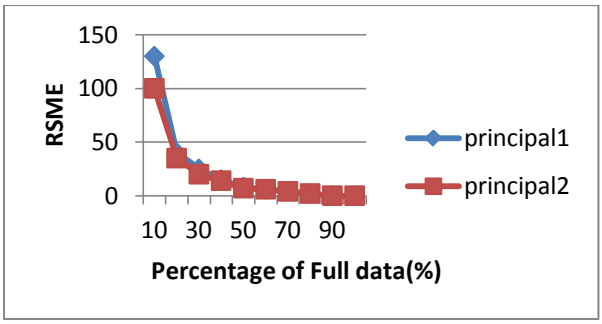


**Figure 3:** Shows Percentage of Full Data vs. Principal 1 and Principal 2

As shown in Figure 3, the results revealed the data percentage with respect to principal 1 and principal 2. The data percentage shows values from 10 to 100 while the principals show values in decreasing trend.

| **Percentage of Full data(%)** | Principal1 | Principal2 |
|---|---|---|
| 10 | 118 | 100 |
| 20 | 45 | 40 |
| 30 | 15 | 16 |
| 40 | 10 | 10 |
| 50 | 5 | 5 |
| 60 | 3 | 3 |
| 70 | 1 | 1 |
| 80 | 0 | 0 |

**Table 2:** Shows Percentage of Full Data vs. Principal 1 and Principal 2

As shown in Table 2, the results revealed the data percentage with respect to principal 1 and principal 2. The data percentage shows values from 10 to 80 while the principals show values in decreasing trend.

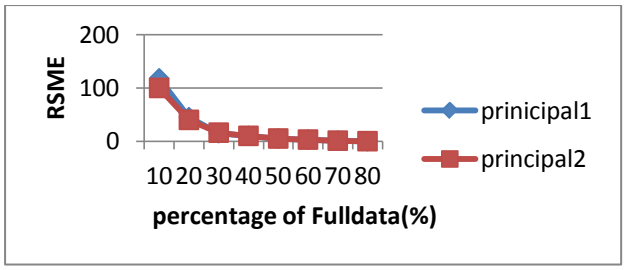


**Figure 4:** Shows Percentage of Full Data vs. Principal 1 and Principal 2

As shown in Figure 4, the results revealed the data percentage with respect to principal 1 and principal 2. The data percentage shows values from 10 to 80 while the principals show values in decreasing trend.

| | EA-RFD-2 | EA-RFD-1 | E-RFD | Ranking-RFD | Rating-RFD | Review-RFD |
|---|---|---|---|---|---|---|
| Tiny Pets | 2.5 | 2.6 | 3.5 | 4 | 6.8 | 8 |
| Social Girl | 4 | 4.1 | 7.3 | 6.5 | 8.5 | 6.7 |
| Fluff Friends | 1.2 | 1.3 | 2.5 | 3.7 | 5.5 | 6.5 |
| Top Girl | 1.5 | 1.7 | 1.6 | 1.8 | 7.3 | 7 |
| VIP Poker | 3 | 3.1 | 5.3 | 5.2 | 4 | 5.5 |
| Sweet shop | 4 | 4.2 | 6 | 8.5 | 6.3 | 8.2 |
| Crime city | 2.8 | 2.9 | 3 | 5.3 | 5 | 9 |

Table 3: Shows Different applications and Fraud Detected

As shown in Table 3, it is evident that different applications and their corresponding fraud probabilities with respect to rating, ranking and review are shown.

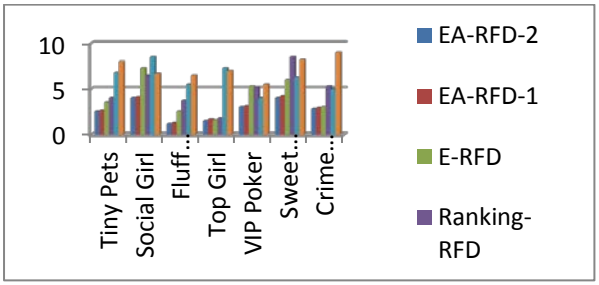


Figure 5: Shows Different applications and Fraud Detected

As shown in Figure 5, it is evident that different applications and their corresponding fraud probabilities with respect to rating, ranking and review are shown.

| **Number of Leading Events(Log)** | Top Free 300 |
|---|---|
| 1 | 500 |
| 2 | 450 |
| 3 | 400 |
| 4 | 350 |
| 5 | 300 |
| 7 | 250 |
| 10 | 200 |
| 15 | 150 |
| 20 | 100 |
| 25 | 50 |
| 30 | 10 |
| 35 | 1 |
| 40 | 0 |
| 50 | 0 |
| 55 | 0 |
| 60 | 50 |
| 70 | 70 |
| 80 | 50 |

Table 4: Leading Sessions vs. Top Free

As shown in Table 4, it is evident that the number of events and the correonsing top free events are presented.

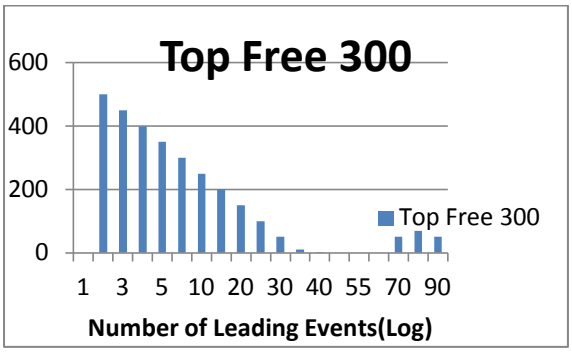


Figure 6: Leading Sessions vs. Top Free

As shown in Figure 6, it is evident that the number of events and the correonsing top free events are presented.

| Number of Leading Session | Top Free 300 |
|---|---|
| 1 | 500 |
| 2 | 400 |
| 3 | 300 |
| 4 | 200 |
| 5 | 100 |
| 6 | 90 |
| 7 | 60 |
| 8 | 50 |
| 9 | 70 |
| 10 | 0 |
| 11 | 50 |

Table 5: Leading Sessions vs. Top Free

As shown in Table 5, it is evident that the number of events and the correonsing top free events are presented.

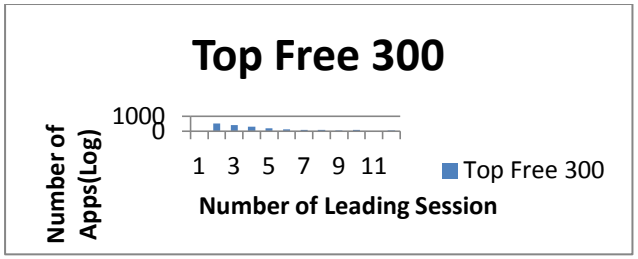


Figure 7: Leading Sessions vs. Top Free

As shown in Figure 7, it is evident that the number of events and the correonsing top free events are presented.

## 4. CONCLUSINOS AND FUTURE WORK

In this paper we studied the problem of ranking fraud for mobile applications. The owners or promoters of mobile apps indulge in bad practices. They often give misleading ranking to their applications in order to attract users. Thus they intend to promote sales. The rationale behind this is to achieve higher revenues. This kind of fraud causes confusion among the genuine users. Therefore it is essential to have a solution for this kind of fraud detection. We proposed a framework in this paper for detecting ranking fraud. We used local and global anomaly approaches. Besides our framework considers three types of evidences from leaing sessions. They are ranking based evidences, review based evidences and rating based evidences. These evidences when aggregated can produce more reliable result that can help in discovering ranking fraud of mobile apps. We built a prototype application that demonstrates the proof of concept besides supporting the local and global anomaly detection as part of finding ranking fraud. The empirical resuts revealed tht the proposed system is useful. We intend to improve the propsed sytem further using data mining algorithms.

## REFERNCES

[1] B. Zhou, J. Pei, and Z. Tang. A spamicity approach to web spam detection. In *Proceedings of the 2008 SIAM International Conferenceon Data Mining*, SDM'08, pages 277–288, 2008.

[2] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly. Detecting spam web pages through content analysis. In *Proceedings of the15th international conference on World Wide Web*, WWW '06, pages 83–92, 2006.

[3] N. Spirin and J. Han. Survey on web spam detection: principles and algorithms. *SIGKDD Explor. Newsl.*, 13(2):50–64, May 2012.

[4] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 939–948, 2010.

[5] Z.Wu, J.Wu, J. Cao, and D. Tao. Hysad: a semi-supervised hybrid shilling attack detector for trustworthy product recommendation. In *Proceedings of the 18th ACM SIGKDD international conference onKnowledge discovery and data mining*, KDD '12, pages 985–993, 2012.

[6] S. Xie, G. Wang, S. Lin, and P. S. Yu. Review spam detection via temporal pattern discovery. In *Proceedings of the 18th ACMSIGKDD international conference on Knowledge discovery and data mining*, KDD '12, pages 823–831, 2012.

[7] B. Yan and G. Chen. Appjoy: personalized mobile application discovery. In *Proceedings of the 9th international conference on Mobilesystems, applications, and services*, MobiSys '11, pages 113–126, 2011.

[8] K. Shi and K. Ali. Getjar mobile application recommendations with very sparse datasets. In *Proceedings of the 18th ACM SIGKDDinternational conference on Knowledge discovery and data mining*, KDD '12, pages 204–212, 2012.

[9] H. Zhu, E. Chen, K. Yu, H. Cao, H. Xiong, and J. Tian. Mining personal context-aware preferences for mobile users. In *DataMining (ICDM), 2012 IEEE 12th International Conference on*, pages 1212–1217, 2012.