

Features Description

This is a description of the features used in the outlier Ensemble detection algorithm to calculate the anomalous scores. All features considered are extracted from promotional reviews.

➤ Given Features:

1. App Features

The following features are computed by aggregating on apps

App_ID: App ID.

App_Reviews: Number of reviews.

App_DUsers: Number of distinct reviews.

App_DCode: Number of distinct codes.

App_DRefferedApps: Number of distinct referred apps.

App_DistinctUser: Number of distinct users.

App_DRefferedApps_per_RefferedApps: Ratio of distinct referred apps to reviews with referred apps.

App_Users: Number of known users.

App_ReferredApps: Number of reviews with referred apps.

App_DCode_Per_Review: Ratio of distinct codes to reviews.

App_DRefferedApps_Per_Rev: Ratio of distinct referred apps to reviews.

App_DUser_Per_Rev: Ratio of distinct users to reviews.

App_DRefferedApp_Per_DUser: Ratio of distinct referred apps to distinct users.

App_Code_Per_DUser: Ratio of distinct codes to distinct users.

App_Avg_Title_length: Average of review title length.

App_Avg_Body_length: Average of review body length.

App_Avg_Rating: Average of review rating.

App_Std_Title_length: Standard deviation of review title length.

App_Std_Body_length: Standard deviation of review body length.

App_Std_Rating: Standard deviation of review rating.

The following three features are aggregated on developers where an app is developed by one developer:

App_Dev_Reviews: Number of reviews for developer.

App_Dev_Apps: Number of promotional apps for developer.

App_Dev_DCode_Per_Reviews: Ratio of distinct code in reviews for developer.

2. User Features

The following features are computed by aggregating on users:

User_ID: User ID.

User_Reviews: Number of reviews.

User_DApps: Number of distinct reviewed apps.

User_DCodes: Number of Distinct codes.

User_DReferredApps: Number of distinct referred apps.

User_DDev: Number of Distinct developer.

User_DReferredApps_Per_ReferredApps: Ratio of distinct referred apps to reviews with referred apps.

User_Apps: Number of apps reviewed.

User_Codes: Number of codes used.

User_ReferredApps: Number of referred Apps.

User_Dev: Number of developer reviewed.

User_DApp_Per_Rev: Ratio of distinct apps in reviews.

User_DCode_Per_Rev: Ratio of distinct code in reviews.

User_DReferredApps_Per_Rev: Ratio of distinct referred app in reviews.

User_DDev_Per_Rev: Number of distinct developer in reviews.

User_DDev_Per_DApps: Ratio of distinct developer in distinct apps.

User_DReferredApp_Per_DApps: Ratio of distinct referred apps in distinct apps.

User_DDev_Per_DReferredApps: Ratio of distinct developer in distinct referred apps.

User_Days: Number of days between first and last review.

User_Avg_Title_Length: Average of review title length.

User_Avg_Body_length: Average of review body length.

User_Avg_Rating: Average of review rating.

User_Std_Title_Length: Standard deviation of review title length.

User_Std_Body_Length: Standard deviation of review body length.

User_Std_Rating_Length: Standard deviation of review rating.

3. Review Features

The following features are computed for each review, without aggregation:

Rev_Text: Review Concatenation of title and body.

Rev_Rating: Review rating.

Rev_LenTitle: Length of review Title.

Rev_LenBody: Length of review Body.

Rev_LenText: Length of review (including body and title).

Rev_Date: Date of review in days.

➤ Novel Features

The following features are computed by aggregating on reviews text, a review text is used in many reviews written by different users and for different apps. This aggregation allows to generate many features for each review text, which is not possible without the proposed aggregation.

Novel_Text: Review text (concatenation of title and body).

Novel_ReferredApp: Referred App.

Novel_AppID: App ID.

Novel_UserID: User ID.

Novel_Avg_Rating: Average of review text rating.

Novel_Std_Rating: Standard deviation of review text rating.

Novel_Title_Length: Length of review title.

Novel_Body_Length: Length of review body.

Novel_Text_Length: Length of review text (including both body and title).

Novel_DDays: Number of distinct dates of review text.

Novel_Days_Per_Rev: Ratio of distinct days in reviews.

Novel_Revs: Number of reviews.

Novel_DistinctApps: Number of distinct apps.

Novel_DApps_Per_Rev: Ratio of distinct apps in reviews.

Novel_Users: Number of known users.

Novel_DUsers: Number of distinct users.

Novel_DUser_Per_Rev: Number of distinct users in reviews.

Novel_DDev: Number of distinct developers.

Novel_DDev_Per_Rev: Ratio of distinct developer in reviews.

Novel_DCodes: Number of distinct codes.

Novel_DCodes_Per_Rev: Ratio of distinct codes in reviews.

Novel_ReferredApps: Number of review referring apps.

Novel_DReferredApps: Number of distinct referred apps.

Novel_DReferredApps_Per_Rev: Ratio of distinct referred apps in reviews.

The following three features are aggregated on codes then aggregated on text review:

Novel_Code_Rev: Number of reviews.

Novel_Code_DRev: Number of distinct text reviews.

Novel_Code_RevRatio: Ratio of distinct text review in reviews.