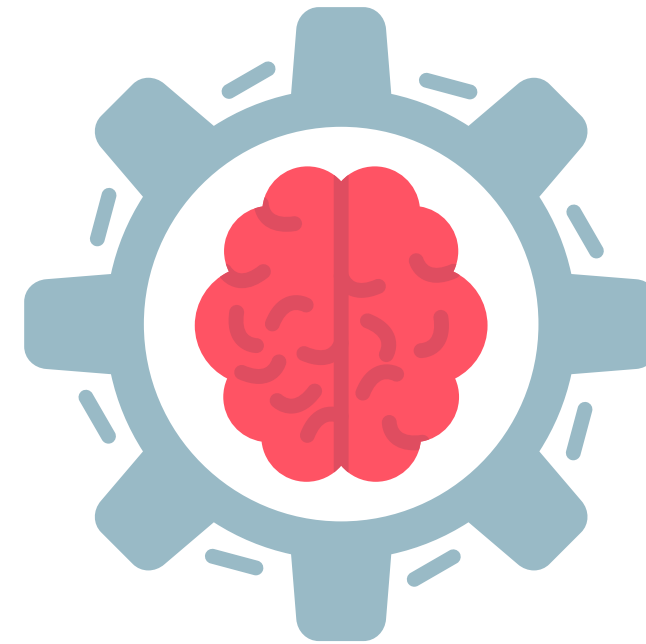


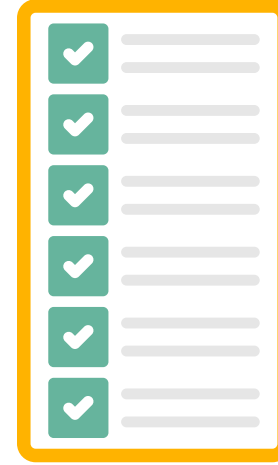
MACHINE LEARNING FOR SOFTWARE ENGINEERING



MANENTI EDOARDO - 0333574

A.A. 2022/2023

AGENDA



1 | INTRODUZIONE

2 | OBIETTIVO

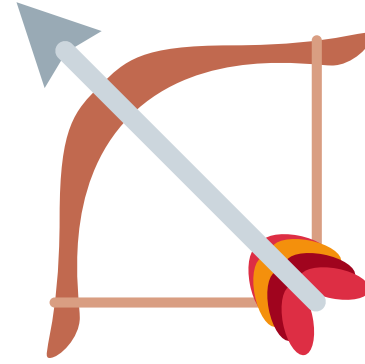
3 | METODOLOGIA

4 | RISULTATI E CONSIDERAZIONI

5 | CONCLUSIONI

6 | LINKS A GITHUB E SONARCLOUD

INTRODUZIONE



NEL MONDO DELL'INGEGNERIA DEL SOFTWARE È FONDAMENTALE:

- GARANTIRE LA **QUALITÀ E L'AFFIDABILITÀ** DI CIÒ CHE VIENE SVILUPPATO.

SOLUZIONE: MECCANISMI DI V&V E QA.

- TENERE IN CONTO IL **BUDGET** E LA GESTIONE DELLE **RISORSE**.

PROBLEMA: COME PREDIRE QUALI CLASSI CONTERRANNO ERRORI?

OBIETTIVO



L'OBIETTIVO GENERALE È SEMPRE STATO CHIARO:

- RENDERE IL PROCESSO DI TESTING PIÙ EFFICIENTE E MIRATO.

NEL CONTESTO DEL PROGETTO "ISW2_CODEMETRICS_PROJECT":

- APPLICARE MODELLI DI **MACHINE LEARNING** PER OTTENERE **PREDIZIONI** SULLA BUGGINESS DELLE CLASSI.
- MOSTRARE L'ANDAMENTO DELLE **PRESTAZIONI** DI TALI MODELLI AL VARIARE DELLE **TECNICHE** APPLICATE SUL DATASET FORNITO PER L'**ADDESTRAMENTO** DEI SUDDETTI.
- IDENTIFICARE LE **CONFIGURAZIONI OTTIMALI** NEI NOSTRI CASI DI ANALISI.

METODOLOGIA



1. RACCOLTA DI **METRICHE** NEL CONTESTO DI 2 PROGETTI OPEN-SOURCE:

- APACHE/BOOKKEEPER
- APACHE/OPENJPA

2. COSTRUZIONE DEI **DATASET** DI TRAINING E TESTING.

3. APPLICAZIONE DI ALGORITMI DI ML PER ESEGUIRE **PREDIZIONI**:

- ANALISI DELLE PRESTAZIONI DI **3 MODELLI**:
 - RANDOM FOREST
 - NAIVE BAYES
 - IBK
- COMBINAZIONE DI **3 TECNICHE**:
 - FEATURE SELECTION
 - BALANCING
 - COST SENSITIVITY

METODOLOGIA

MISURAZIONI 1



IL PROCESSO DI MISURAZIONE È STATO EFFETTUATO ATTRAVERSO DUE STRUMENTI PRINCIPALI:



JIRA: ISSUE TRACKING SYSTEM PER LA RACCOLTA DEI TICKET E DELLE VERSIONI



GIT: VERSION CONTROL SYSTEM PER LA RACCOLTA DEI COMMIT

NB:] NEL PROGETTO "ISW2_CODEMETRICS_PROJECT" TALI OGGETTI RECUPERATI ATTRAVERSO GLI STRUMENTI CITATI SONO STATI ELABORATI IN MODO TALE DA RACCOGLIERE LE INFORMAZIONI UTILI ED ESTRAPOLARNE METRICHE SCRITTE POI SU FILE .CSV E .ARFF PER COSTRUIRE EVENTUALMENTE I DATASETS DI CUI I MODELLI DI ML HANNO BISOGNO.

METODOLOGIA

MISURAZIONI 2



LE AV NON SONO SEMPRE PRESENTI O COMPLETE SUI TICKET
ESTRATTI DA JIRA:

- SI RICORRE AL CALCOLO DELLE AV ATTRAVERSO LA TECNICA DI **PROPORTION INCREMENTAL** SULLA MEDIA DEL PROPORTION DEI TICKET A DISPOSIZIONE FINO A QUEL MOMENTO.
 - QUALORA NON CE NE FOSSERO ABBASTANZA (VIENE IMPOSTATA UNA **TRESHOLD** DI '5') SI RICORRE ALL'APPROCCIO **COLD-START** PRENDENDO LA MEDIANA TRA I VALORI MEDI DEL PROPORTION CALCOLATI SUI TICKET CONSISTENTI (CON INFORMAZIONI SULLE AV) DI **ALTRI PROGETTI APACHE** RITENUTI SIGNIFICATIVI PER ESEGUIRE QUESTO TIPO STIMA.

$$p = \frac{FV - IV}{FV - OV} \quad IV = \max\{1, FV - (FV - OV) \cdot p\} \quad \left(\text{se } FV = OV \rightarrow p = \frac{FV - IV}{1} \right)$$

METODOLOGIA

METRICHE 1



| | |
|------------------------|------------------------------------------------------------------------------------------------|
| SIZE | Dimensione in LOC della classe. |
| LOC_ADDED* | Somma delle LOC aggiunte. |
| LOC_REMOVED* | Somma delle LOC rimosse. |
| LOC_TOUCHED* | Somma LOC_ADDED + LOC_REMOVED. |
| CHURN* | Fattore di aggiunta e rimozione di LOC, pari a $ \text{LOC_ADDED} - \text{LOC_REMOVED} $. |
| NUMBER_OF_REVISIONS | Numero di revisioni della classe. |
| NUMBER_OF_DEFECT_FIXES | Numero di revisioni della classe che risolvono bug issues di JIRA. |
| NUMBER_OF_AUTHORS | Numero di autori. |
| IS_BUGGY | Metrica binaria relativa alla bugginess della classe nella release considerata. |

NB]: DELLE METRICHE
SEGNATE CON '*' SONO
STATI CALCOLATI ANCHE
VALORI *_MAX E *_AVG
MA NON SONO STATI
INSERITI IN TABELLA
PER LEGGIBILITÀ E
RIDONDANZA.

METODOLOGIA

METRICHE 2



- TUTTE LE METRICHE INTRODOTTE SONO DA CONSIDERARSI **INTRA-RELEASE** E NON CUMULATIVE.
- LA PRIMA METÀ DELLE METRICHE È MOLTO AFFINE E RIGUARDA LE **LOC**, QUESTA SCELTA È STATA FATTA ANCHE PER OSSERVARE COME AGISCE LA **TECNICA DI FEATURE SELECTION** IN QUESTI CASI.
- L'ULTIMA METRICA RELATIVA ALLA **BUGGINESS** VIENE CALCOLATA SECONDO IL SEGUENTE CRITERIO:
 - VENGONO PRESI TUTTI I **TICKET** DI TIPO "BUG" CON RISOLUZIONE "FIXED" E CON STATO "CLOSED O "RESOLVED" DA JIRA.
 - OGNI CLASSE JAVA MODIFICATA DA UN COMMIT CHE È LINKATO AD UNO DEI TICKET ESTRATTI VIENE **ETICHETTATA** COME BUGGY PER TUTTE LE **AFFECTED VERSIONS**.

METODOLOGIA

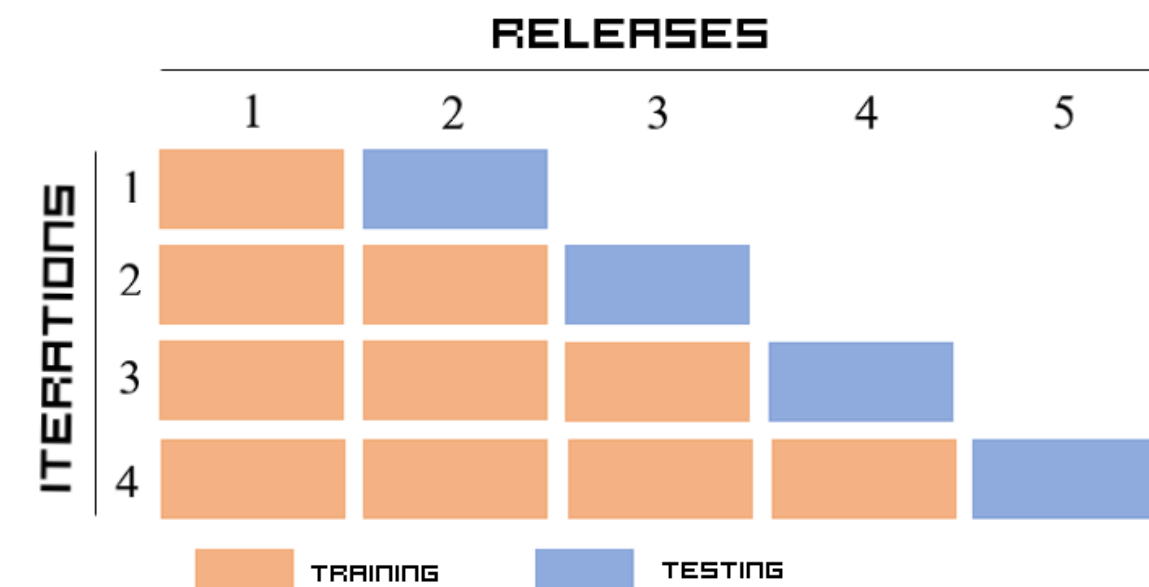
DATASETS



- PER EFFETTUARE LE PREDIZIONI, I CLASSIFICATORI NECESSITANO DI ESSERE ADDESTRATI ATTRAVERSO **TRAINING DATASETS**.
- PER RACCOGLIERE LE VALUTAZIONI SUI CLASSIFICATORI, ABBIAMO BISOGNO DI VERIFICARE LE PREDIZIONI ATTRAVERSO I **TESTING DATASETS**.

PER COSTRUIRE I SET UTILIZZIAMO LA TECNICA DI **WALK FORWARD**, CHE È DI TIPO **TIME-SERIES** ED È UN **PROCESSO ITERATIVO**:

1. COSTRUZIONE DEL TRAINING SET CON LE PRIME 'K' RELEASES (VIENE RIESEGUITO IL LABELING ESCLUSIVAMENTE IN BASE ALLE **INFORMAZIONI DISPONIBILI FINO A QUEL MOMENTO**).
2. COSTRUZIONE DEL TESTING SET CHE PER OGNI ITERAZIONE CONTERRÀ LE INFORMAZIONI SULLE CLASSI DELLA RELEASE K+1-ESIMA (SU CUI ANDRANNO FATTE LE PREDIZIONI) CHE AVRANNO IL LABELING IN BASE A **TUTTE LE INFORMAZIONI DISPONIBILI**.



METODOLOGIA

PREDIZIONI



PER COME ABBIAMO COSTRUITO I DATASET:

- IL TRAINING SET SI DIRÀ **AFFETTO DA SNORING** E SU DI ESSO VERRANNO APPLICATE LE VARIE TECNICHE DI SELECTION E BALANCING.
- MENTRE **LO SNORING SARÀ ASSENTE** NEL TESTING SET CHE SFRUTTA TUTTI I DATI DI CUI SIAMO A CONOSCENZA E SU DI ESSO NON VIENE APPLICATA ALCUNA TRASFORMAZIONE.

NB]: PER MITIGARE GLI EFFETTI DELLO SNORING È STATO SCARTATO IL 50% DELLE RELEASE PIÙ RECENTI DAL DATASET, PRIMA CHE QUESTO FOSSE VALUTATO DAL CLASSIFICATORE.

UNA VOLTA COSTRUITI I **FILE .ARFF** CONTENENTI I SET, PER APPLICARE LE TECNICHE DI FEATURE SELECTION E BALANCING E POI **ESEGUIRE LE VALUTAZIONI** DELLE **PREDIZIONI** FATTE DAI CLASSIFICATORI (CON UNA CERTA COMBINAZIONE DI TECNICHE) ABBIAMO SFRUTTATO L'**API DI WEKA** EFFETTUANDO POI UN RISCONTRO CON LA **GUI DI WEKA**.



RISULTATI



LE METRICHE DI INTERESSE SONO:

- **PRECISION**
- **RECALL**
- **KAPPA**
- **AREA UNDER ROC (AUC)**

LE VARIANTI PER LE TECNICHE DI BALANCING SONO:

- **OVERSAMPLING**
- **UNDERSAMPLING**
- **SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE)**

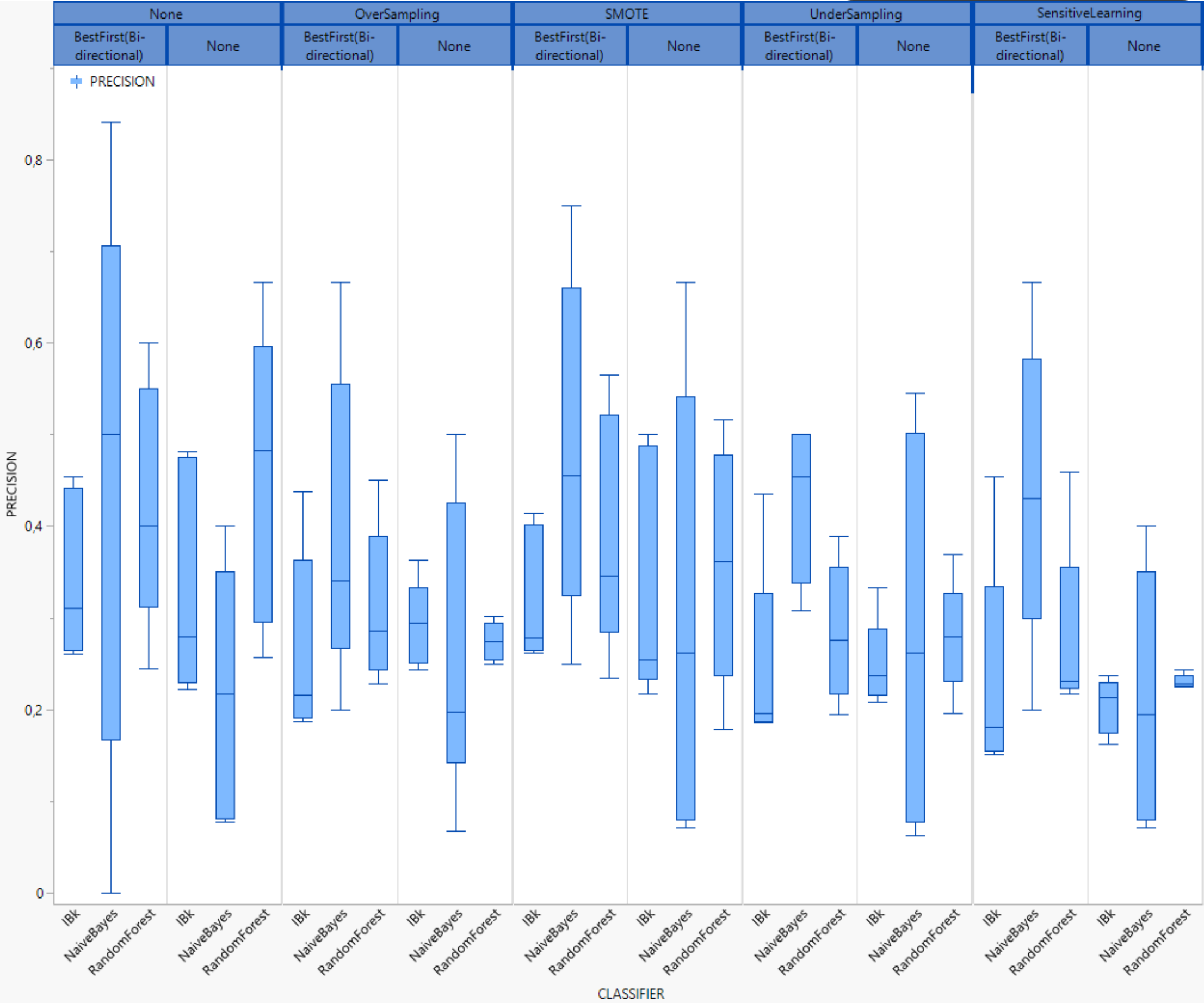
LA TECNICA DI SELECTION USATA È INVECE LA **BEST FIRST BI-DIRECTIONAL**

PER ANALIZZARE LE METRICHE OTTENUTE DALLA VALUTAZIONE DI WEKA ED INSERITE IN DEI FILE **.CSV** È STATO UTILIZZATO "JMP" UN PROGRAMMA DI ESPLORAZIONE STATISTICA CHE CI HA PERMESSO DI REALIZZARE DEI **GRAFICI** CON CUI ESPORRE I **RISULTATI**.



PRECISION

BOOKKEEPER

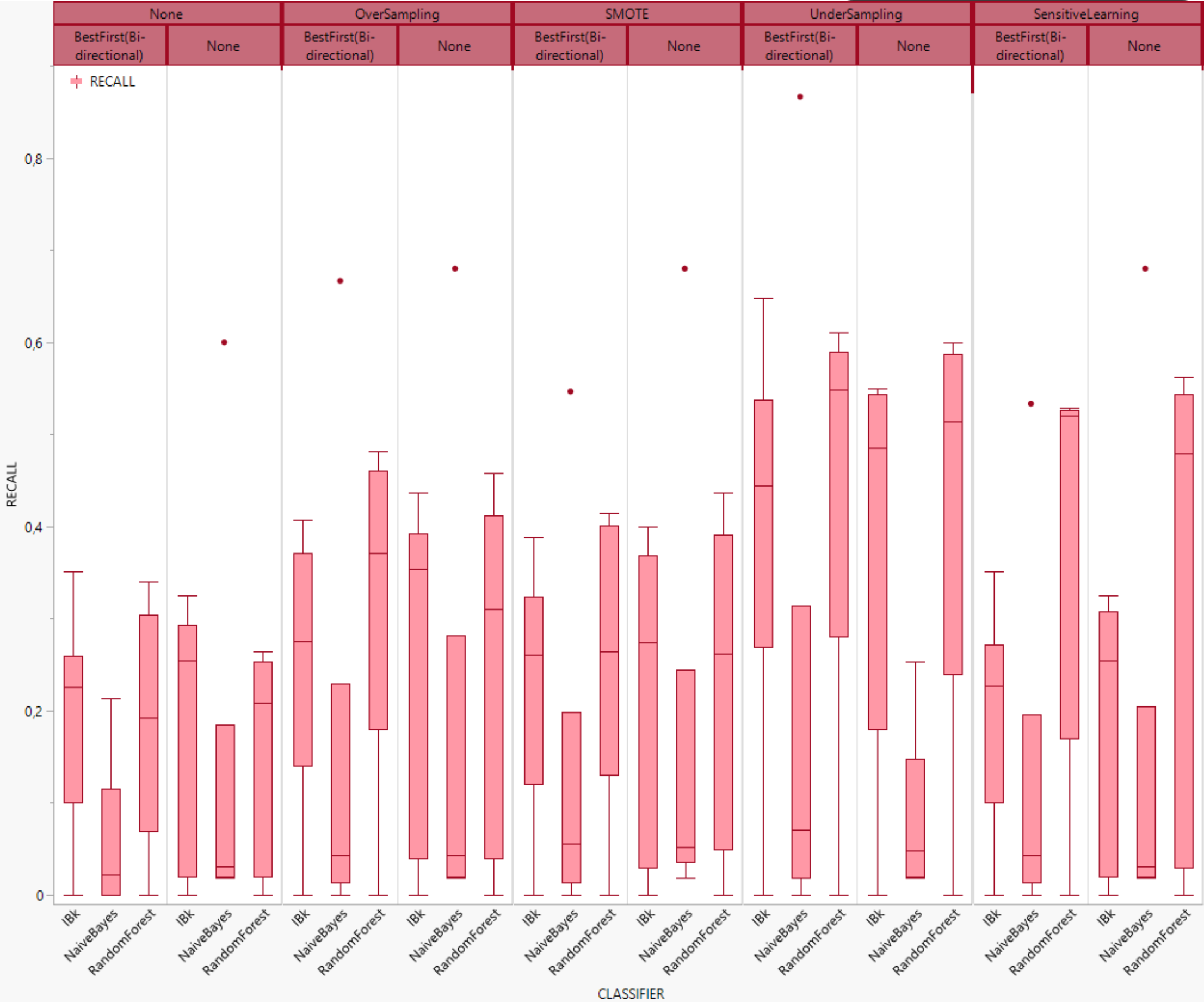


- **FEATURE SELECTION** MIGLIORA SIGNIFICATIVAMENTE LA PRECISION DEL CLASSIFICATORE NAIVE BAYES E LIEVEMENTE QUELLA DI IBK. TUTTAVIA, PEGGIORA QUELLA DI RANDOM FOREST.
- **OVERSAMPLING** PEGGIORA LA PRECISION DI TUTTI I CLASSIFICATORI TRANNE IBK, MA APPLICANDO **ANCHE FEATURE SELECTION** SI REGISTRA UN MIGLIORAMENTO SIGNIFICATIVO SULLA PRECISION DI NAIVE BAYES, MENTRE IN IBK VIENE PEGGIORATA.
- **SMOTE** NON SEMBRA IMPATTARE MOLTO LA PRECISION DEI CLASSIFICATORI QUANDO USATA DA SOLA, APPLICANDO **ANCHE FEATURE SELECTION** INVECE C'È UN AUMENTO DELLA PRECISION IN NAIVE BAYES.
- **UNDERSAMPLING** SEMBRA PEGGIORARE MOLTO LA PRECISION IN RANDOM FOREST SE USATA DA SOLA, MENTRE SEMBRA MIGLIORARLA IN NAIVE BAYES QUANDO **COMBINATA CON LA FEATURE SELECTION**.
- **SENSITIVE LEARNING** PEGGIORA LA PRECISION IN TUTTI I CLASSIFICATORI. IN **COMBINAZIONE CON FEATURE SELECTION** SI OTTIENE UN MIGLIORAMENTO DELLA PRECISION IN NAIVE BAYES MENTRE SU IBK E RANDOM FOREST IL PEGGIORAMENTO RIMANE

RECALL

BOOKKEEPER

MANENTI EDOARDO - 0333574 - A.A. 2022/2023

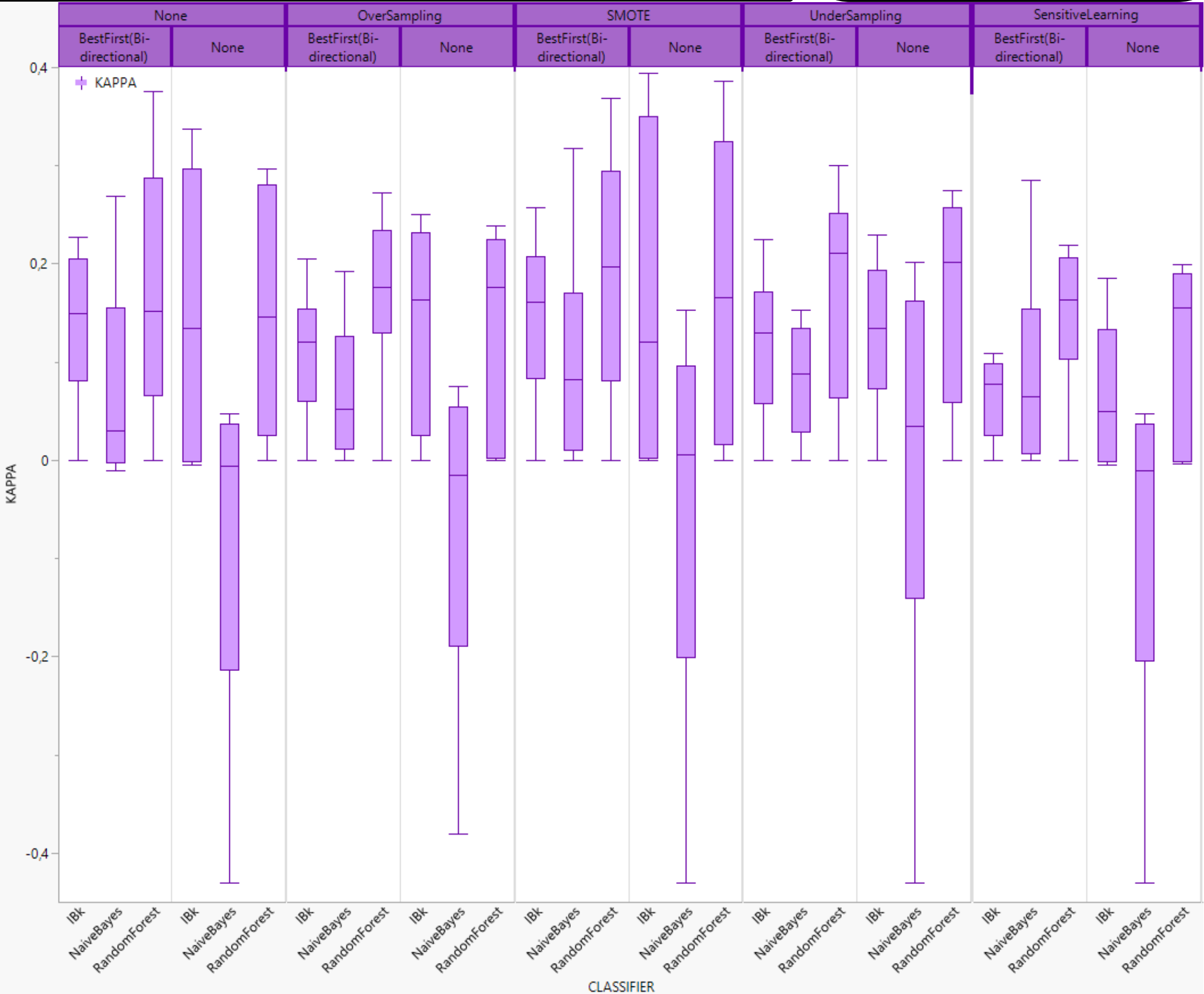


- **FEATURE SELECTION** NON SEMBRA AVERE ALCUNA INFLUENZA SULLA RECALL DEI CLASSIFICATORI.
- **OVERSAMPLING** AUMENTA SIGNIFICATIVAMENTE LA RECALL IN IBK E RANDOM FOREST. INOLTRE, APPLICANDOLA **INSIEME A FEATURE SELECTION** LA RECALL AUMENTA ANCORA IN RANDOM FOREST MA DIMINUISCE LIEVEMENTE IN IBK.
- **SMOTE** AUMENTA LIEVEMENTE LA RECALL IN IBK E RANDOM FOREST. I RISULTATI OSSERVATI SONO SIMILI ANCHE CON L'**AGGIUNTA DI FEATURE SELECTION**.
- **UNDERSAMPLING** HA FINORA L'IMPATTO PIÙ SIGNIFICATIVO SULLA RECALL FACENDO RAGGIUNGERE I VALORI MASSIMI A IBK E RANDOM FOREST. QUEST'ULTIMO CLASSIFICATORE CONTINUA A MIGLIORARE QUANDO TALE TECNICA VIENE **COMBINATA CON LA FEATURE SELECTION** MENTRE SI REGISTRA UN LIEVE PEGGIORAMENTO IN IBK.
- **SENSITIVE LEARNING** AUMENTA DI MOLTO LA RECALL IN RANDOM FOREST. IN **COMBINAZIONE CON FEATURE SELECTION** SI OTTIENE UN ULTERIORE MIGLIORAMENTO NELLO STESSO CLASSIFICATORE MENTRE RIMANGONO IN TUTTI E DUE I CASI PRESSOCHÈ INVARIATI I VALORI DI RECALL DEGLI ALTRI CLASSIFICATORI.

KAPPA

BOOKKEEPER

MANENTI EDOARDO - 0333574 - A.A. 2022/2023

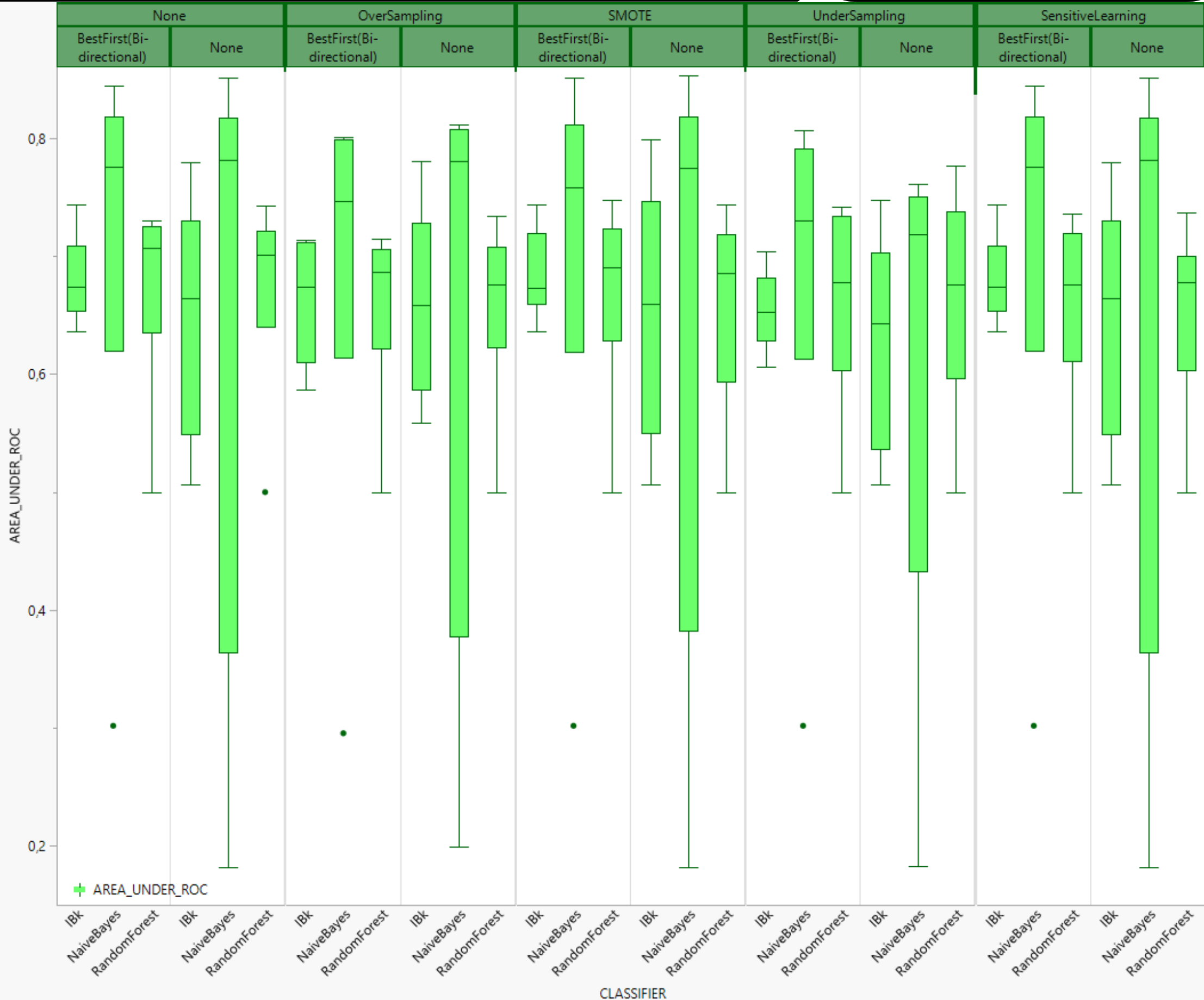


- **FEATURE SELECTION** MIGLIORA LA KAPPA DEL CLASSIFICATORE NAIVE BAYES E MOLTO LIEVEMENTE QUELLA DEGLI ALTRI DUE CLASSIFICATORI.
- **OVERSAMPLING** MIGLIORA LIEVEMENTE LA KAPPA DI TUTTI I CLASSIFICATORI TRANNE NAIVE BAYES, CHE INVECE REGISTRA UN MIGLIORAMENTO QUANDO SI CONSIDERA **ANCHE FEATURE SELECTION** CHE PERÒ PEGGIORA LA KAPPA DI IBK.
- **SMOTE** MIGLIORA LA KAPPA DI RANDOM FOREST, APPLICANDO **ANCHE FEATURE SELECTION** INVECE C'È UN AUMENTO SIGNIFICATIVO DELLA KAPPA IN TUTTI I CLASSIFICATORI.
- **UNDERSAMPLING** SEMBRA MIGLIORARE I VALORI DI KAPPA IN TUTTI I CLASSIFICATORI. IN PARTICOLARE QUESTO ACCADE QUANDO **COMBINATA CON LA FEATURE SELECTION** TRANNE PER IBK CHE IN QUESTO CASO MANTIENE I VALORI PRECEDENTI.
- **SENSITIVE LEARNING** PEGGIORA LA KAPPA IN IBK E MIGLIORA QUELLA IN RANDOM FOREST. IN **COMBINAZIONE CON FEATURE SELECTION** SI OTTIENE UN MIGLIORAMENTO SIGNIFICATIVO DELLA KAPPA IN NAIVE BAYES E LIEVEMENTE IN IBK.

AUC

BOOKKEEPER

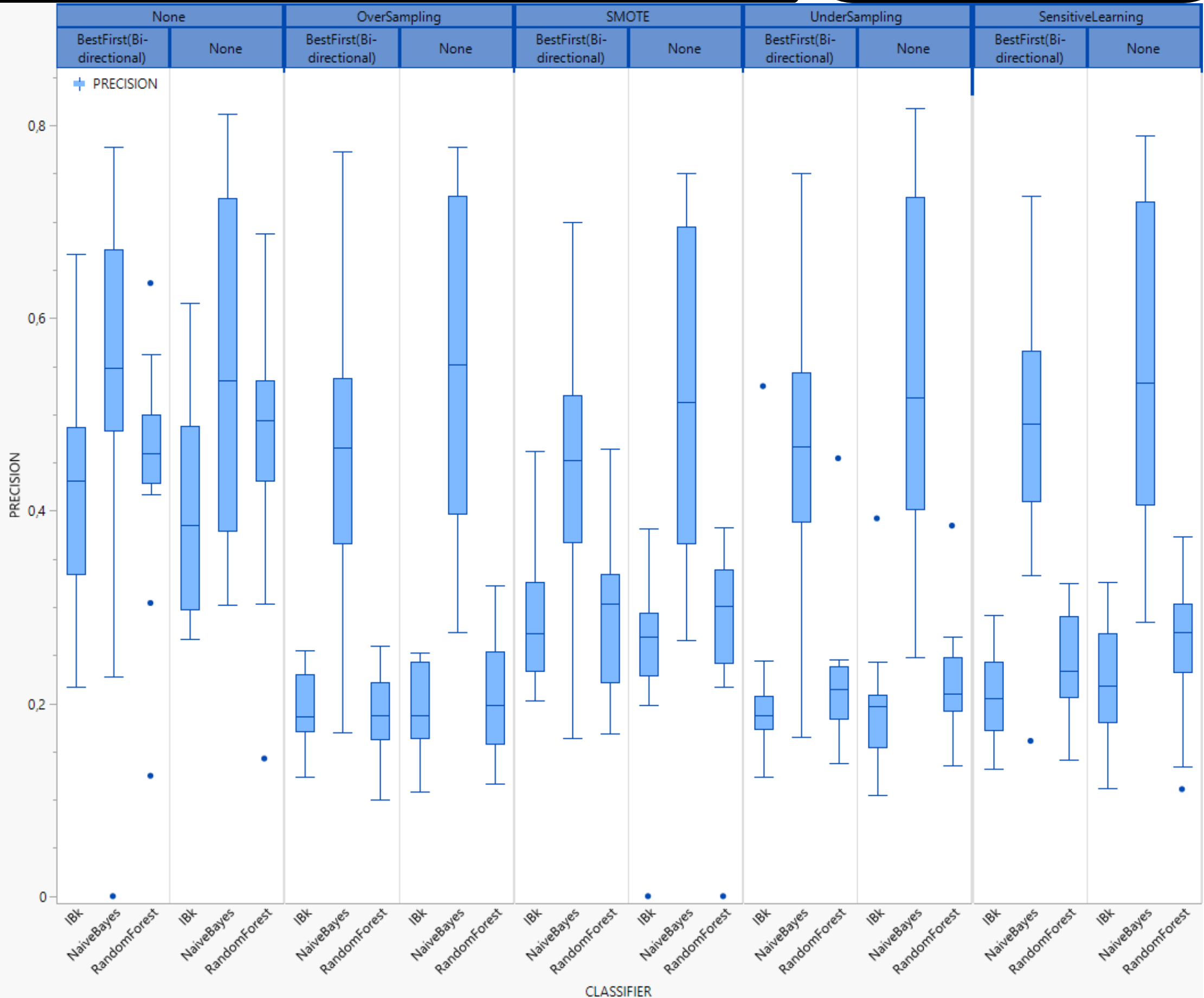
MANENTI EDOARDO - 0333574 - A.A. 2022/2023



- **FEATURE SELECTION** RIDUCE LA VARIABILITÀ NEI VALORI DI AUC REGISTRATI DAI VARI CLASSIFICATORI.
- **OVERSAMPLING** PEGGIORA LIEVEMENTE LA AUC DI IBK E RANDOM FOREST. L'**AGGIUNTA DI FEATURE SELECTION** MIGLIORA L'AUC NEI DUE CLASSIFICATORI MENZIONATI MA LA PEGGIORA PER NAIVE BAYES.
- **SMOTE** NON IMPATTA MOLTO LA AUC QUANDO APPLICATA DA SOLA, APPLICANDO **ANCHE FEATURE SELECTION** INVECE C'È UN AUMENTO LIEVE DELLA AUC IN TUTTI I CLASSIFICATORI TRANNE NAIVE BAYES.
- **UNDERSAMPLING** SEMBRA PEGGIORARE I VALORI DI AUC IN TUTTI I CLASSIFICATORI. TALE EFFETTO VIENE MITIGATO INVECE QUANDO **COMBINATA CON LA FEATURE SELECTION**.
- **SENSITIVE LEARNING** NON SEMBRA AVERE ALCUN IMPATTO SULLA AUC REGISTRATA DAI CLASSIFICATORI. IN **COMBINAZIONE CON FEATURE SELECTION** SI OTTIENE UNA RIDUZIONE DELLA VARIABILITÀ DEI VALORI REGISTRATI DA IBK E NAIVE BAYES.

PRECISION

OPENJPA



- **FEATURE SELECTION** MIGLIORA LA PRECISION DI IBK E LIEVEMENTE QUELLA DI NAIVE BAYES. TUTTAVIA, PEGGIORA QUELLA DI RANDOM FOREST.

- **OVERSAMPLING** PEGGIORA NOTEVOLMENTE LA PRECISION DI TUTTI I CLASSIFICATORI TRANNE NAIVE BAYES E APPLICANDO **INOLTRE FEATURE SELECTION** SI REGISTRA UN PEGGIORAMENTO ANCHE IN QUEST'ULTIMO CLASSIFICATORE.

- **SMOTE** PEGGIORA LA PRECISION SOPRATTUTTO IN IBK E RANDOM FOREST, COMBINANDO **ANCHE FEATURE SELECTION** TALE PEGGIORAMENTO È MITIGATO TRANNE CHE IN NAIVE BAYES CHE IN QUESTO CASO REGISTRA UN CALO DI PRECISION.

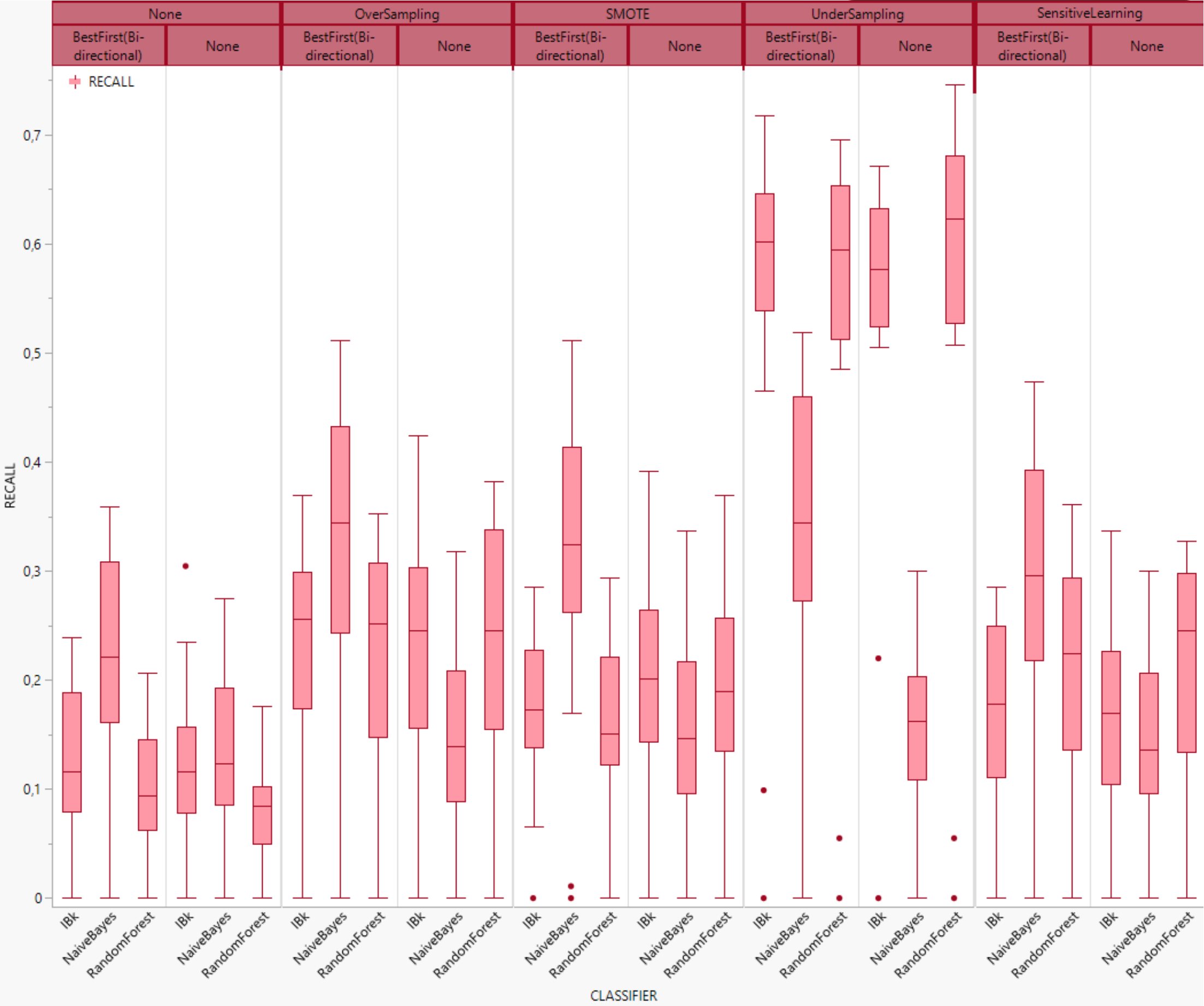
- **UNDERSAMPLING** PEGGIORA MOLTO LA PRECISION DI IBK E RANDOM FOREST, CON L'**AGGIUNTA DI FEATURE SELECTION** PEGGIORA ANCHE NAIVE BAYES.

- **SENSITIVE LEARNING** REGISTRA GLI STESSI EFFETTI DEI DUE PRECEDENTI.

SI VIENE AD IDENTIFICARE UN PATTERN DI PEGGIORAMENTO PER TUTTE LE TECNICHE TRANNE LA FEATURE SELECTION IN CUI PERÒ I MIGLIORAMENTI NON SONO SIGNIFICATIVI.

RECALL

OPENJPA



- **FEATURE SELECTION** AUMENTA LA RECALL DI NAIVE BAYES E LEGGERMENTE QUELLA DI RANDOM FOREST.
- **OVERSAMPLING** AUMENTA SIGNIFICATIVAMENTE LA RECALL IN IBK E RANDOM FOREST. INOLTRE, APPLICANDOLA **INSIEME A FEATURE SELECTION** LA RECALL AUMENTA ANCHE PER NAIVE BAYES CHE CON QUESTA COMBINAZIONE SUPERA I VALORI REGISTRATI DAGLI ALTRI DUE CLASSIFICATORI.
- **SMOTE** AUMENTA LA RECALL DI TUTTI I CLASSIFICATORI. TUTTAVIA, CON L'**AGGIUNTA DI FEATURE SELECTION** L'AUMENTO DIMINUISCE PER IBK E RANDOM FOREST MENTRE AUMENTA SIGNIFICATIVAMENTE PER NAIVE BAYES.
- **UNDERSAMPLING** HA NUOVAMENTE L'IMPATTO PIÙ SIGNIFICATIVO SULLA RECALL FACENDO RAGGIUNGERE I VALORI MASSIMI A IBK E RANDOM FOREST. PER NAIVE BAYES INVECE IL MIGLIORAMENTO SI HA QUANDO TALE TECNICA VIENE **COMBINATA CON LA FEATURE SELECTION**.
- **SENSITIVE LEARNING** AUMENTA LA RECALL IN IBK E RANDOM FOREST. IN **COMBINAZIONE CON FEATURE SELECTION** SI OTTIENE UN MIGLIORAMENTO SIGNIFICATIVO INVECE IN NAIVE BAYES.

KAPPA

OPENJPA

MANENTI EDOARDO - 0333574 - A.A. 2022/2023

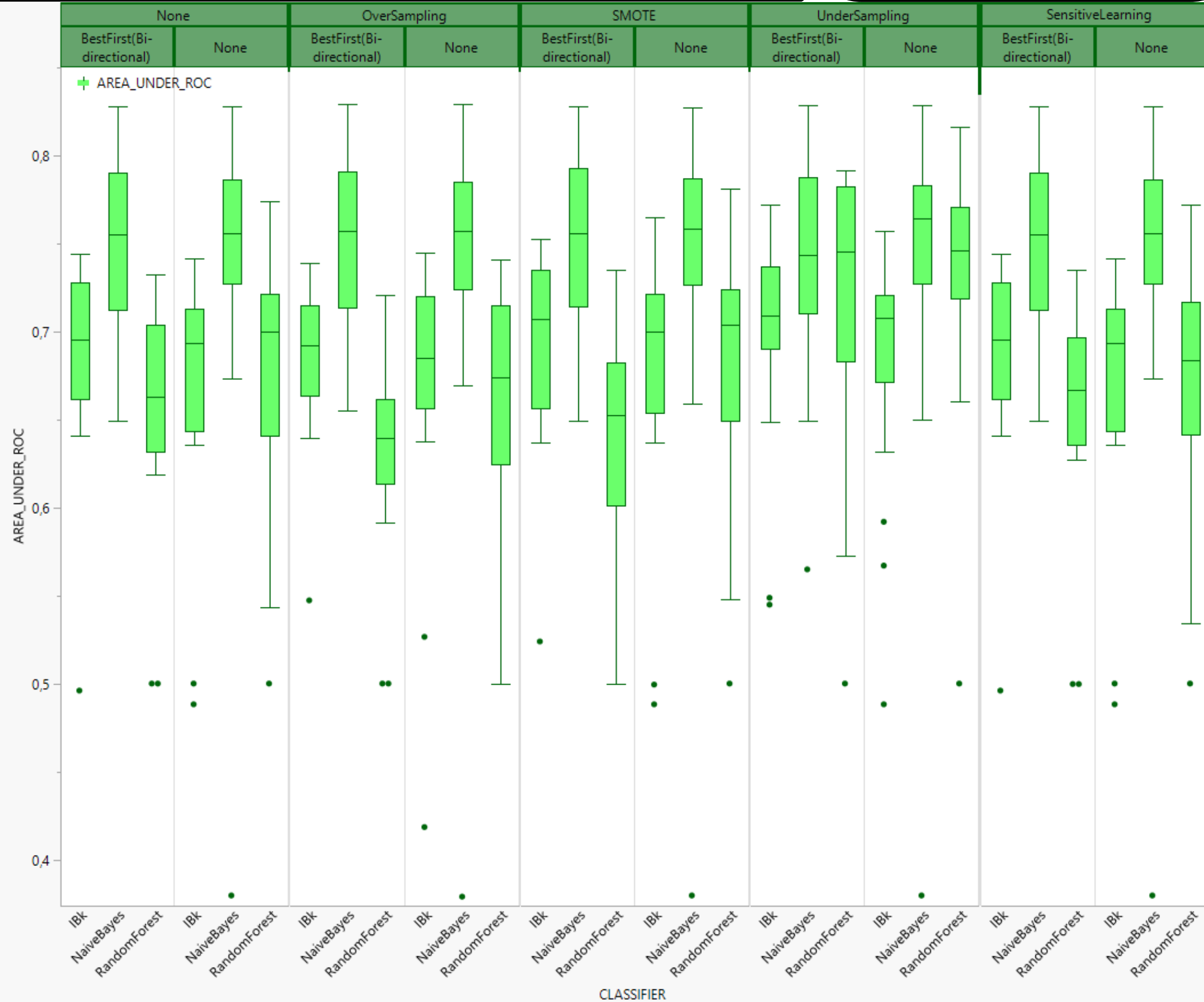


- **FEATURE SELECTION** MIGLIORA LA KAPPA DEL CLASSIFICATORE NAIVE BAYES E MOLTO LIEVEMENTE QUELLA DI IBK.
- **OVERSAMPLING** PEGGIORA LA KAPPA IN IBK MA MIGLIORA LIEVEMENTE LA KAPPA DI RANDOM FOREST E NAIVE BAYES, CHE POI REGISTRA UN ENORME MIGLIORAMENTO QUANDO SI CONSIDERA **ANCHE FEATURE SELECTION**.
- **SMOTE** MIGLIORA LA KAPPA DI TUTTI I CLASSIFICATORI E APPLICANDO **ANCHE FEATURE SELECTION** C'È UN AUMENTO SIGNIFICATIVO DELLA KAPPA IN NAIVE BAYES.
- **UNDERSAMPLING** SEMBRA MIGLIORARE I VALORI DI KAPPA IN TUTTI I CLASSIFICATORI. IN PARTICOLARE QUESTO ACCADE IN NAIVE BAYES QUANDO QUESTA TECNICA VIENE **COMBINATA CON LA FEATURE SELECTION**.
- **SENSITIVE LEARNING** PEGGIORA LA KAPPA IN IBK E MIGLIORA LIEVEMENTE QUELLA IN RANDOM FOREST E NAIVE BAYES. IN **COMBINAZIONE CON FEATURE SELECTION** SI OTTIENE UN MIGLIORAMENTO SIGNIFICATIVO DELLA KAPPA IN NAIVE BAYES E LIEVEMENTE IN IBK.

AUC

OPENJPA

MANENTI EDOARDO - 0333574 - A.A. 2022/2023



- **FEATURE SELECTION** DIMINUISCE I VALORI DI AUC IN RANDOM FOREST.
- **OVERSAMPLING** PEGGIORA LA AUC DI IBK E RANDOM FOREST. L'**AGGIUNTA DI FEATURE SELECTION** PEGGIORA ULTERIORMENTE L'AUC IN RANDOM FOREST.
- **SMOTE** NON IMPATTA MOLTO LA AUC QUANDO APPLICATA DA SOLA, APPLICANDO **ANCHE FEATURE SELECTION** INVECE C'È UN AUMENTO LIEVE DELLA AUC IN TUTTI I CLASSIFICATORI TRANNE RANDOM FOREST CHE PEGGIORA SIGNIFICATIVAMENTE.
- **UNDERSAMPLING** SEMBRA AUMENTARE I VALORI DI AUC IN TUTTI I CLASSIFICATORI. INVECE, QUANDO **COMBINATA CON LA FEATURE SELECTION** QUESTA TECNICA DIMINUISCE I VALORI DI AUC REGISTRATI PER NAIVE BAYES.
- **SENSITIVE LEARNING** DIMINUISCE LA AUC DI RANDOM FOREST. TALE RIDUZIONE VIENE AMPLIFICATA APPLICANDO LA TECNICA IN **COMBINAZIONE CON FEATURE SELECTION**.

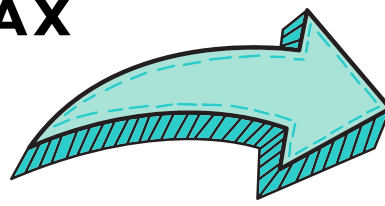
CONCLUSIONI

SELECTION

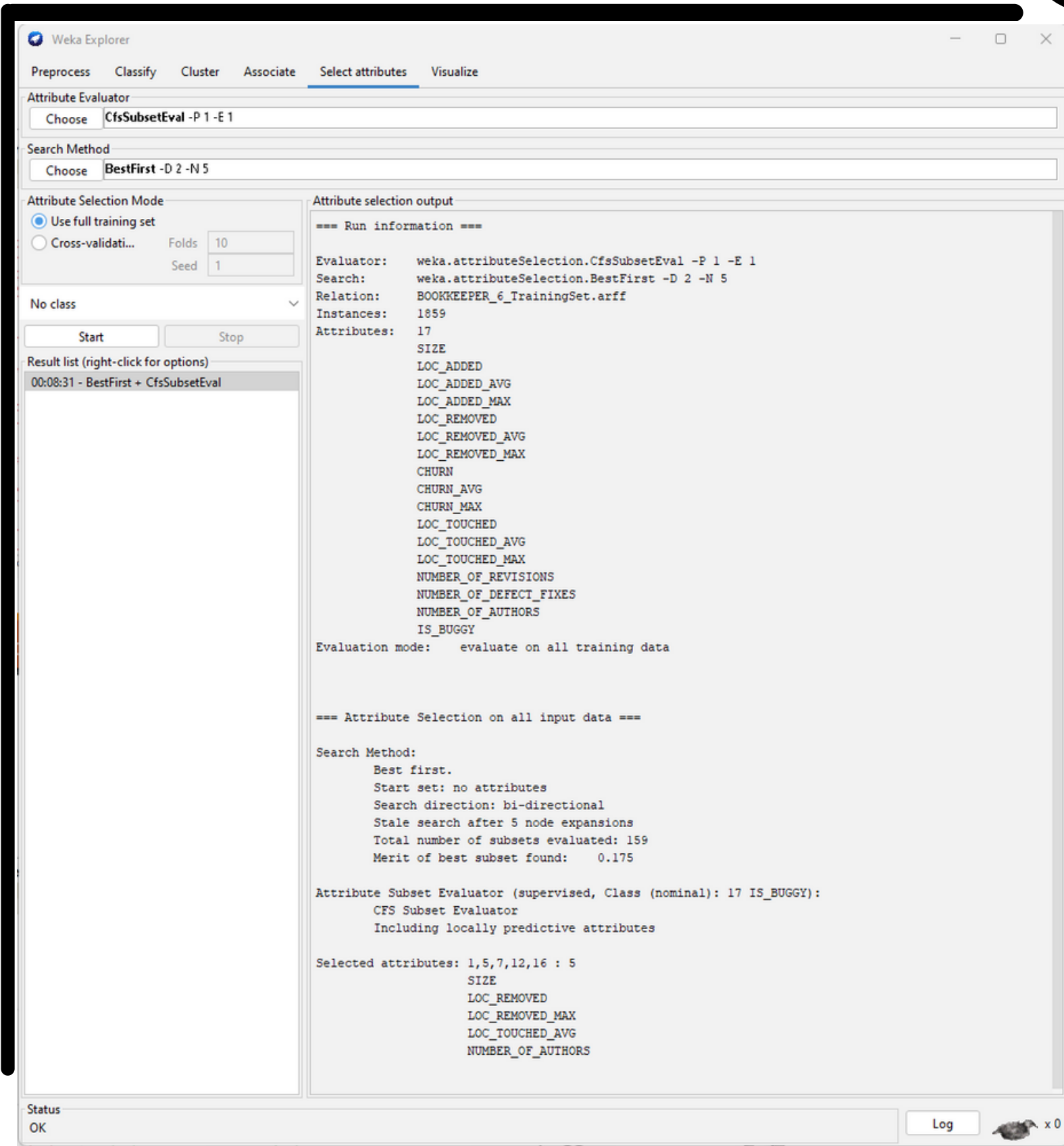


APPLICANDO LA FEATURE SELECTION È POSSIBILE INDIVIDUARE 6 TRA I 17 ATTRIBUTI INIZIALI CHE HANNO UNA RILEVANZA MAGGIORE RISPETTO ALLA VARIABILE DI CLASSIFICAZIONE ("IS_BUGGY"):

- SIZE
- ~~LOC_ADDED~~
- ~~LOC_ADDED_AVG~~
- ~~LOC_ADDED_MAX~~
- ~~LOC_REMOVED~~
- ~~LOC_REMOVED_AVG~~
- ~~LOC_REMOVED_MAX~~
- ~~CHURN~~
- ~~CHURN_AVG~~
- ~~CHURN_MAX~~
- ~~LOC_TOUCHED~~
- ~~LOC_TOUCHED_AVG~~
- ~~LOC_TOUCHED_MAX~~
- ~~NUMBER_OF_REVISIONS~~
- ~~NUMBER_OF_DEFECT_FIXES~~
- NUMBER_OF_AUTHORS



- SIZE ✓
- LOC_REMOVED ✓
- LOC_REMOVED_MAX ✓
- LOC_TOUCHED_AVG ✓
- NUMBER_OF_AUTHORS ✓



CONCLUSIONI

BALANCING



TRA LE TECNICHE DI BILANCIAMENTO, NON CE N'È UNA CHE È NETTAMENTE MIGLIORE DELLE ALTRE, MA DIPENDE INTRINSECAMENTE DAL DATASET DI PARTENZA E DAL CLASSIFICATORE CONSIDERATO.

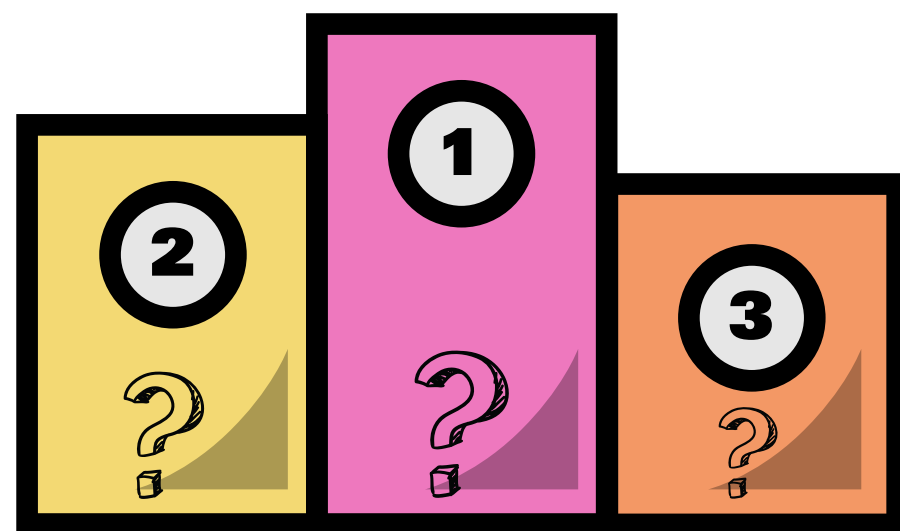


TUTTAVIA POSSIAMO INDIVIDUARE **DUE TENDENZE** INTERESSANTI:

1. **IL SUCCESSO**, SU ENTRAMBI I PROGETTI, **DELLA TECNICA DI UNDERSAMPLING** PER L'AUMENTO DELLA **RECALL** SU TUTTI I CLASSIFICATORI.
2. **IMPATTO MAGGIORE**, IN MEDIA, SU **OPENJPA** DELLE TECNICHE DI **BALANCING** (DOVUTO EVIDENTEMENTE AD UN **MAGGIOR SBILANCIAMENTO** DI OPENJPA)

CONCLUSIONI

CLASSIFIERS



TRA I DIVERSI CLASSIFICATORI, NON CE N'È UNO CHE HA PERFORMANCE MIGLIORI DEGLI ALTRI... ANCHE IN QUESTO CASO I RISULTATI VARIANO IN BASE AL DATASET INIZIALE E ALLE VARIE TECNICHE APPLICATE.

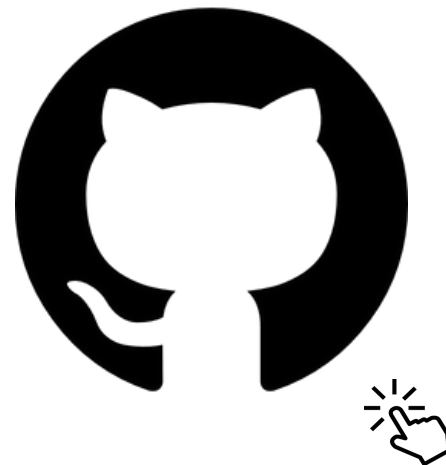
IN GENERALE IL **CLASSIFICATORE CON PERFORMANCE PIÙ BASSE** RISULTA ESSERE **NAIVE BAYES** CHE TUTTA VIA MANTIENE BUONE PERFORMANCE SULLA PRECISION RISPETTO A IBK E RANDOM FOREST.

INOLTRE DAL MOMENTO CHE È STATO USATO **WALKFORWARD** COME TECNICA DI VALIDAZIONE, È POSSIBILE CHE CI SIANO **ITERAZIONI PARTICOLARMENTE SFORTUNATE CHE PESANO NEGATIVAMENTE SULLE PRESTAZIONI**. GENERANDO MOLTEPLICI ALBERI DI DECISIONE SU CUI MEDIARE IL RISULTATO FINALE DELLE PRESTAZIONI, RANDOMFOREST RIESCE A MITIGARE I RISULTATI NEGATIVI AUMENTANDO LE PERFORMANCE GENERALI.

GRAZIE PER L'ATTENZIONE!



LINKS A GITHUB E SONARCLOUD



[HTTPS://GITHUB.COM/EDOMAN000/ISW2_CODEMETRICS_PROJECT](https://github.com/EDOMAN000/ISW2_CODEMETRICS_PROJECT)



[HTTPS://SONARCLOUD.IO/PROJECT/OVERVIEW?ID=EDOMAN000_ISW2_CODEMETRICS_PROJECT](https://sonarcloud.io/project/overview?id=EDOMAN000_ISW2_CODEMETRICS_PROJECT)