EXPERIMENT REPORT

Student Name	Sahil Kotak
Project Name	36120 Advanced Machine Learning Application Spring 2023 - Assignment 1
Date	01-09-2023
Deliverables	• kotak_sahil-24707592-week3_ensemble_model_5.ipynb • Logistic Regression & SVM Stacking Model

---

1. EXPERIMENT BACKGROUND	
Provide information about the problem/project such as the scope, the overall objective, expectations. Lay down the goal of this experiment and what are the insights, answers you want to gain or level of performance you are expecting to reach.	
1.a. Business Objective	This project aims to predict which college basketball players are likely to be drafted into the NBA based on their statistics from the current season of college basketball. This prediction can be used by NBA teams to make informed decisions during the draft process, by sports analysts to provide insights into the draft prospects, and by the players themselves to gauge their likelihood of being drafted. Accurate results will improve draft decisions, increase team performance, and more insightful draft analysis. Inaccurate results may lead to suboptimal draft decisions, decreased team performance, and misleading draft analysis. The stakes are high, as a team's success in the NBA can have significant financial implications, including higher ticket sales, increased merchandise sales, and more lucrative sponsorship deals. Therefore, developing a reliable and accurate predictive model for this purpose is crucial.
1.b. Hypothesis	Our central hypothesis for this week was that optimizing the feature space by including engineered features and focusing on feature importance would improve the performance of our ensemble model without leading to overfitting. **Objective:** The primary objective was to enhance the model's predictive power while maintaining its generalization capability. We introduced several new features based on domain knowledge and statistical insights, such as anomaly scores and reevaluated our approach of dropping features like ‘pick’. We also implemented advanced ensemble techniques like ElasticNet and cost-sensitive training. **Rationale:** **Interpretability and Feature Importance**: While re-evaluating the importance of the features in our dataset, we realized that *‘pick’* has a prediction power of 0.985, and that makes it our most valuable feature, considering that we will find a better way to handle the data preparation for the feature.

	Robustness through Anomaly Detection: The main rationale for generating anomaly scores was to capture the "outlier" nature inherent in the draft pick selection process. In the context of NBA drafts, being selected is somewhat analogous to being an "anomaly" within a sea of talented players. Anomaly scores help the model to identify these unique, draft-worthy talents by quantifying how much each player deviates from the norm, thereby improving the model's ability to predict such rare events more accurately.
1.c. Experiment Objective	The primary objective of this experiment is to investigate whether the integration of anomaly detection scores and better feature engineering combined with fine-tuned ensemble methods can further improve the predictive performance of our models. Specifically, we aim to achieve an AUROC score above our previous best, which was 0.9913, while maintaining model interpretability and avoiding overfitting. **Possible Scenarios:** ● **Successful Outcome:** The ensemble model incorporating new data preparation and feature engineering achieves an AUROC score greater than 0.9913 and generalizes well to the test set, validating our approach. ● **Partial Success:** The ensemble model improves performance but at the cost of overfitting, indicating a need for further regularization or complexity reduction. ● **No Improvement:** The ensemble model does not show any meaningful improvement in AUROC, suggesting that the anomaly scores and/or additional ensemble complexity do not contribute to predictive power for this particular problem. ● **Degradation:** The model performs worse than simpler models, suggesting that the added complexity and features could be introducing noise rather than improving performance.

---

2. EXPERIMENT DETAILS	
Elaborate on the approach taken for this experiment. List the different steps/techniques used and explain the rationale for choosing them.	
2.a. Data Preparation	Along with the steps from Week 1 & 2 experiments, the following additional steps were taken: ● **Handling 'pick' and 'Rec_Rank' Columns:** These columns were transformed into binary features. Any missing values were set to 0, and all other values were set to 1. **Rationale:** ‘pick’ and ‘Rec_Rank’ had 97.53% and 69.63% missing values respectively. Initially, we decided to drop both the features, but given their importance when compared with the target variable we decided better handling of both is required. The exact pick number or recruiting rank may not be as informative as whether a player was picked or ranked at all. Converting these to binary features simplifies the model while preserving this essential information. ● **Handling 'dunks_ratio' Column:** Missing values in 'dunks_ratio' were calculated using the 'dunksmade' and 'dunksmiss_dunksmade' columns wherever possible. **Rationale:** The ratio of dunks made to dunks attempted is a valuable feature

	but it had 54.9% missing values. Calculating this ratio using related columns maximizes the use of available data. ● **Handling 'mid_ratio' and 'rim_ratio' Columns:** Missing values were calculated using the 'midmade', 'midmade_midmiss', 'rimmade', and 'rimmade_rimmiss' columns. **Rationale:** Like 'dunks_ratio', these ratios can be valuable for understanding a player's effectiveness but had missing values. Again, we used related columns to impute these missing values. **Steps Not Taken:** ● **Feature Elimination:** Though we initially experimented with feature importance techniques to reduce dimensionality, we ultimately decided against it as it didn't yield improved performance.
2.b. Feature Engineering	**Steps Taken:-** ● **Anomaly Scores:** Anomaly scores were generated for each row in the dataset using the Isolation Forest algorithm. **Rationale:** Our problem of predicting draft picks is akin to finding anomalies in the data. Players who get drafted are exceptional in one way or another, making them outliers or 'anomalies' in the data. The anomaly score feature aims to capture this aspect quantitatively. ● **Encoding and Scaling:** Categorical variables were one-hot encoded, and numerical features were scaled using Standard Scaler, consistent with previous experiments. ● **Features Removed:** None. We retained all the original features while adding the anomaly scores for this experiment. **Future Considerations:** **Interaction Features:** In future experiments, we might consider adding interaction terms among features that are highly correlated with the target variable.
2.c. Modelling	**Models Trained:** ● **Ensemble Techniques:** Enhanced last week's stacking model to include ElasticNet and cost-sensitive Logistic Regression. ● **Hyperparameters Tuned:** ● **ElasticNet:** Focused on alpha and l1_ratio. ● **Cost-sensitive Training:** Experimented with different class weights for Logistic Regression. ● **Grid Search:** Fine-tuned the ensemble's meta-model.

	Models Not Trained: Avoided complex models like Random Forest and XGBoost due to the dataset's linear nature and strong performance of simpler models. **Future Considerations:** May explore different meta-models for stacking and advanced hyperparameter optimization techniques. **Highlight for Future Experiments:** The continued superior AUROC performance of the Stacking Ensemble model, even this week, continues to reaffirm our hypothesis about the dataset's predominantly linear nature. This insight is crucial for directing future modeling strategies and experiments.

---

3. EXPERIMENT RESULTS	
Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified.	
3.a. Technical Performance	**Models Trained:** **Stacking Ensemble:** Extended the previous week's stacking model that had shown promising results. Added ElasticNet and cost-sensitive Logistic Regression to the base models. **Voting Classifier:** Included as a simpler ensemble method for comparison. **Hyperparameters Tuned:** • **ElasticNet:** Tuned alpha and l1_ratio to optimize regularization. • **Cost-sensitive Logistic Regression:** Adjusted class weights to handle imbalanced classes. • **Grid Search:** Refined the meta-model's hyperparameters for the stacking ensemble. **Models Not Trained:** Opted not to experiment with non-linear models like Random Forests or XGBoost due to the linear nature of the dataset and the strong performance of simpler models. **Scores:** **Stacking Ensemble:** 0.99789 **Stacking Ensemble with Anomaly Feature:** 0.99933 **Voting Classifier:** 0.99075

<table>
<tr><td></td><td>

Future Considerations:

Hyperparameter optimization techniques like Randomized Search could be explored to improve the stacking ensemble further potentially.

In this experiment, we built upon our successful stacking ensemble model from last week, focusing on refining and optimizing it. The addition of an anomaly feature improved the AUROC score, demonstrating its potential utility in future experiments.

We also added ElasticNet and cost-sensitive Logistic Regression to the ensemble, fine-tuning them through grid search. The strong performance validates our focus on linear models and ensemble techniques.

</td></tr>
<tr><td>

3.b. Business Impact

</td><td>

Referencing the previous week's insights, the experiments in the current week continue to underscore the model's potential as a powerful predictive tool for NBA draft decisions. The high AUROC scores, especially from the Stacking and Ensemble models, further instill confidence in our predictive capabilities.

However, business implications go beyond just model performance. The minute differences in AUROC between the models, although statistically significant, need to be weighed against practical considerations. For instance, the added complexity of the Stacking model over the SVM might not justify its marginally better performance in a real-world setting.

</td></tr>
<tr><td>

3.c. Encountered Issues

</td><td>

- **Missing Value Imputation for Special Columns:** Columns like 'Rec_Rank', 'pick', 'dunks_ratio', and 'mid/rim ratio' required special handling for missing values.

  Solution: Custom strategies like ratio computation and placeholder values were implemented for each column.

This week's experiments also brought their unique challenges, particularly around missing value imputation for specialized columns and convergence issues for logistic models. Solutions were devised for these, and they paved the way for more focused challenges in future experiments, particularly around fine-tuning and avoiding overfitting.

For Future Experiments:

- **Overfitting Concerns:** With an AUROC score close to 1, there's a risk of model overfitting. For the Future: Utilizing techniques like cross-validation or introducing regularization terms to avoid this.

</td></tr>
</table>

---

4. FUTURE EXPERIMENT
Reflect on the experiment and highlight the key information/insights you gained from it that are valuable for the overall project objectives from a technical and business perspective.

4.a. Key Learning	**Anomaly Detection as a Feature:** The addition of an anomaly score feature significantly boosted our model's performance, affirming that the task of predicting draft picks is akin to detecting anomalies in the data. **Refinement over Complexity:** Adding an ElasticNet model and employing cost-sensitive training in our stacking ensemble did not show a significant improvement over the previous week's simpler models. This reinforces the notion that adding complexity needs to be carefully weighed against the potential performance gains. **Fine-Tuning Importance:** Grid search allowed us to fine-tune our models, but the gains were marginal. This suggests that our models are already quite optimized for the problem at hand and that hyperparameter tuning might offer diminishing returns. **Future Direction:** Given the current high performance and the risk of overfitting, further experimentation should focus on model refinement rather than complexity. Techniques to prevent overfitting, such as regularization and cross-validation, could be beneficial. Given the success of the anomaly score, additional feature engineering may also yield fruitful results.
4.b. Suggestions / Recommendations	Given the results achieved and the overall objective of the project, potential that we want to explore: **Further Feature Engineering:** - Rationale: The anomaly score was a significant addition. Additional feature engineering, such as interaction terms or polynomial features, could capture more complex relationships in the data. - Expected Uplift: Moderate to High. Given the large impact of the anomaly score, other feature engineering efforts might also yield significant gains. **Focus on Advanced Ensemble Techniques:** - Rationale: Given that the stacking ensemble with the anomaly feature provided the highest score, exploring more advanced ensemble techniques could be the logical next step. Techniques like weighted ensembling, where each model's prediction is weighted based on its performance, or adaptive boosting could be explored to enhance the predictive power further. - Expected Uplift: Moderate to High. Since the ensemble of simpler models has already shown promise, more sophisticated ensembling methods have a good chance of squeezing out additional performance gains. This could make the model more robust and improve its generalizability to unseen data.