**Data set 2: Best model:** Stacking-Best ML models-PCA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **RMSE** | **MAE** | **MAD** | **MAPE** | **Best Params** | **Tuning Method** |
| Linear regression | 0.476 | 0.32 | 0.32 | 1.52 | intercept:0.26328492612607945  Estiamted coefficients:[ 0.93567656 0.10150702 -0.05991236 -0.03259918 0.06483268 -0.09183363  0.03179132 0.07053715 0.00097763 -0.04941848 0.00978014 0.02217405  0.07414503 -0.13984545 -0.02671262 0.14877269 -0.13336105 0.07468454  -0.04294315 0.04604471 -0.01134979 -0.0512252 0.03037838 -0.01892212  0.07008345 -0.11491878 0.12315587 0.10673824 -0.17838002 0.02866013] |  |
| Linear regression-Optimized | 0.476 | 0.32 | 0.32 | 1.52 | Best Parameters: {'fit\_intercept': True}  Coefficients: [ 0.93567656 0.10150702 -0.05991236 -0.03259918 0.06483268 -0.09183363  0.03179132 0.07053715 0.00097763 -0.04941848 0.00978014 0.02217405  0.07414503 -0.13984545 -0.02671262 0.14877269 -0.13336105 0.07468454  -0.04294315 0.04604471 -0.01134979 -0.0512252 0.03037838 -0.01892212  0.07008345 -0.11491878 0.12315587 0.10673824 -0.17838002 0.02866013]  Intercept: 0.26328492612607945 | Random Search 'fit\_intercept': [True, False] |
| Polynomial Regression -PCA | 0.82 | 0.64 | 0.57 | 3.07 | Best Polynomial degree: 5 |  |
| Polynomial Regression-Optimized- PCA | 0.48 | 0.33 | 0.23 | 1.58 | {'pca\_\_n\_components': 12, 'poly\_\_degree': 1} | Random Search 'poly\_\_degree': randint(1, 6) |
| Gaussian process | 1.203 | 0.709 | 0.404 | 3.063 | kernel = RBF(length\_scale=1.0) |  |
| Gaussian process-Optimized-PCA | 0.63 | 0.45 | 0.35 | 2.08 | {'kernel\_\_k1\_\_constant\_value': 8.424426408004217, 'kernel\_\_k2\_\_length\_scale': 2.2233911067827616} | Random Search  'kernel\_\_k1\_\_constant\_value': uniform(0.1, 10.0),      'kernel\_\_k2\_\_length\_scale': uniform(0.1, 10.0) |
| MA-Uni | 3.23 | 2.60 | 2.60 | 12.82 | window\_size = 5 |  |
| MA-Uni-Optimized | 2.536 | 2.063 | 2.063 | 10.024 | Best Window Size: 2 | Random Search  window\_sizes = range(2, 21) |
| MA-Multi | 3.38 | 2.72 | 2.72 | 13.46 | window\_size = 5 |  |
| MA-Multi-Optimized | 3.07 | 2.51 | 2.51 | 12.25 | Best Window Size: 3 | Random Search  window\_sizes = range(2, 21) |
| Arima-Uni | 21.67 | 21.36 | 21.36 | 97.34 | p, d, q = 5, 1, 0 |  |
| Arima-Uni-Optimized | 21.659 | 21.353 | 21.353 | 97.290 | Best ARIMA(4, 1, 3) | Random Search  p\_range = range(0, 6)  d\_range = range(0, 3)  q\_range = range(0, 6) |
| Arima-Multi | 0.53 | 0.39 | 0.39 | 1.82 | SARIMAX(4,2,5) |  |
| Arima-Multi-Optimized | 0.474 | 0.321 | 0.321 | 1.502 | Best SARIMAX(3, 2, 1) | Random Search  p\_range = range(0, 6)  d\_range = range(0, 3)  q\_range = range(0, 6) |
| Stacking- With best STmodels | RMSE: 0.47 | MAE: 0.32 | MAD: 0.21 | R-squared: 0.98 | ('gp', gp\_model),  ('poly', poly\_model),  ('sarimax', sarimax\_model),  ('lr', lr\_model) | kernel = C(8.424426408004217, (1e-2, 1e2)) \* RBF(2.2233911067827616, (1e-2, 1e2))  gp\_model = GaussianProcessRegressor(kernel=kernel, n\_restarts\_optimizer=9)  sarimax\_model = CustomSARIMAX(order=(3, 2, 1))  poly\_model = CustomPolynomialRegression(degree=1, n\_components=12)  lr\_model = LinearRegression() |
| Stacking- With best ST models-PCA | RMSE: 0.46 | MAE: 0.32 | MAD: 0.21 | R-squared: 0.98 | ('gp', gp\_model),          ('poly', poly\_model),          ('sarimax', sarimax\_model),          ('lr', lr\_model)  'pca\_\_n\_components': 12 | kernel = C(8.424426408004217, (1e-2, 1e2)) \* RBF(2.2233911067827616, (1e-2, 1e2))  gp\_model = GaussianProcessRegressor(kernel=kernel, n\_restarts\_optimizer=9)  sarimax\_model = CustomSARIMAX(order=(3, 2, 1))  poly\_model = CustomPolynomialRegression(degree=1, n\_components=12)  lr\_model = LinearRegression() |
| KNN | 0.51 | 0.35 | 0.25 | 1.66 | Number of Neighbors: 3 | n\_neighbors = [3, 5, 7, 9, 11, 13, 15, 17] |
| KNN-Optimized | 0.45 | 0.30 | 0.21 | None | Best parameters found: {'n\_neighbors': 2, 'p': 2, 'weights': 'distance'} | Random Search  param\_distributions = {      'n\_neighbors': randint(1, 31),      'weights': ['uniform', 'distance'],      'p': [1, 2]  } |
| Decision Tree | 0.50 | 0.36 | 0.27 | 1.67 | Max Depth: 5 | max\_depths = [3, 5, 7, 9, 11, 13, 15, 17] |
| Decision Tree-Optimized | 0.49 | 0.34 | 0.24 | 1.60 | Best parameters found: {'max\_depth': 5, 'min\_samples\_leaf': 9, 'min\_samples\_split': 6} | Random Search  param\_dist = {      'max\_depth': randint(3, 20),      'min\_samples\_split': randint(2, 20),      'min\_samples\_leaf': randint(1, 20)  } |
| SVR | 0.50 | 0.36 | 0.26 | 1.71 |  | kernel='rbf', C=1.0, epsilon=0.1 |
| SVR-Optimized | 0.45 | 0.31 | 0.21 | 1.48 | Best parameters found: {'C': 88.62326508576255, 'epsilon': 0.020736445177905044, 'gamma': 0.00039459088111000007} | Random Search  param\_dist = {      'C': loguniform(1e-2, 1e2),      'gamma': loguniform(1e-4, 1e-1),      'epsilon': loguniform(1e-4, 1e-1)  } |
| Stacking-Best ML models | 0.4225 | 0.2931 | 0.2221 | -1.3688% |  | # Decision Tree with optimized parameters  decision\_tree = DecisionTreeRegressor(max\_depth=5, min\_samples\_leaf=9, min\_samples\_split=6)  # KNN with optimized parameters  knn = KNeighborsRegressor(n\_neighbors=2, p=2, weights='distance')  # SVR with optimized parameters  svr = make\_pipeline(StandardScaler(), SVR(C=88.62326508576255, epsilon=0.020736445177905044, gamma=0.00039459088111000007)) |
| Stacking-Best ML models-PCA | 0.4137 | 0.2824 | 0.2067 | -1.3394% | ('knn', knn),          ('decision\_tree', decision\_tree),          ('svr', svr)  'pca\_\_n\_components': 12 | # Decision Tree with optimized parameters  decision\_tree = DecisionTreeRegressor(max\_depth=5, min\_samples\_leaf=9, min\_samples\_split=6)  # KNN with optimized parameters  knn = KNeighborsRegressor(n\_neighbors=2, p=2, weights='distance')  # SVR with optimized parameters  svr = make\_pipeline(StandardScaler(), SVR(C=88.62326508576255, epsilon=0.020736445177905044, gamma=0.00039459088111000007))  n\_components = 12 |
| GRU | 3.6411 | 3.1925 | 3.1900 | 15.4293 | Lag : 24  Dropout(rate=0.1)  First GRU layer: 200 units  Second GRU layer: 100 units  Third GRU layer: 50 units  Fourth GRU layer: 50 units |  |
| GRU-Optimized | 3.6146 | 3.0961 | 3.0990 | 15.2954 | Lag : 24,{'model\_\_gru\_units4': 50, 'model\_\_gru\_units3': 25, 'model\_\_gru\_units2': 50, 'model\_\_gru\_units1': 200, 'model\_\_dropout\_rate': 0.3} | Random Search      'model\_\_gru\_units1': [100, 200, 300],      'model\_\_gru\_units2': [50, 100, 150],      'model\_\_gru\_units3': [25, 50, 75],      'model\_\_gru\_units4': [25, 50],      'model\_\_dropout\_rate': [0.1, 0.2, 0.3, 0.4, 0.5] |
| LSTM | 3.6684 | 3.2411 | 3.2407 | 15.5286 | Lag : 24  Dropout(rate=0.1)  First LSTM layer: 200 units  Second LSTM layer: 100 units  Third LSTM layer: 50 units  Fourth (last) LSTM layer: 50 units |  |
| LSTM-Optimized | 3.6313 | 3.1664 | 3.1660 | 15.3869 | Lag : 24,{'model\_\_lstm\_units4': 25, 'model\_\_lstm\_units3': 25, 'model\_\_lstm\_units2': 50, 'model\_\_lstm\_units1': 200, 'model\_\_dropout\_rate': 0.1} | Random Search      'model\_\_lstm\_units1': [100, 200],      'model\_\_lstm\_units2': [50, 100],      'model\_\_lstm\_units3': [25, 50, 75],      'model\_\_lstm\_units4': [25, 50],      'model\_\_dropout\_rate': [0.1, 0.2] |
| BILSTM | 3.6224 | 3.1451 | 3.1510 | 15.3639 | Lag : 24  Dropout(rate=0.1)  First BILSTM layer: 200 units  Second BILSTM layer: 100 units  Third BILSTM layer: 50 units  Fourth (last) BILSTM layer: 50 units |  |
| BILSTM-Optimized | 3.6142 | 3.0973 | 3.1019 | 15.2994 | Lag : 24,{'model\_\_dropout\_rate': 0.2, 'model\_\_bilstm\_units4': 50, 'model\_\_bilstm\_units3': 25, 'model\_\_bilstm\_units2': 50, 'model\_\_bilstm\_units1': 200} | Random Search      'model\_\_bilstm\_units1': [100, 200],      'model\_\_bilstm\_units2': [50, 100],      'model\_\_bilstm\_units3': [25, 50, 75],      'model\_\_bilstm\_units4': [25, 50],      'model\_\_dropout\_rate': [0.1, 0.2] |
| BIGRU | 3.6098 | 3.0446 | 3.0712 | 15.2985 | Lag : 24  Dropout(rate=0.1)  First BIGRU layer: 200 units  Second BIGRU layer: 100 units  Third BIGRU layer: 50 units  Fourth (last) BIGRU layer: 50 units |  |
| BIGRU-Optimized | 3.5980 | 3.0885 | 3.1159 | 15.3449 | Lag : 24,{'model\_\_dropout\_rate': 0.1, 'model\_\_bigru\_units4': 25, 'model\_\_bigru\_units3': 50, 'model\_\_bigru\_units2': 50, 'model\_\_bigru\_units1': 100} | Random Search      'model\_\_bigru\_units1': [100, 200],      'model\_\_bigru\_units2': [50, 100],      'model\_\_bigru\_units3': [25, 50, 75],      'model\_\_bigru\_units4': [25, 50],      'model\_\_dropout\_rate': [0.1, 0.2] |
| CNN | 4.1015 | 3.3577 | 3.3214 | 16.5387 | Lag : 24  Filters1=64, kernel\_size=3, activation='relu'  Dropout(0.1)  Filters2=128, kernel\_size=3, activation='relu'  Dropout(0.1)  Filters3=128, kernel\_size=3, activation='relu'  MaxPooling1D(pool\_size=2) |  |
| CNN-Optimized | 3.9690 | 3.2711 | 3.2745 | 16.2689 | Lag : 24,{'model\_\_kernel\_size2': 3, 'model\_\_kernel\_size1': 5, 'model\_\_filters2': 256, 'model\_\_filters1': 32, 'model\_\_dropout\_rate': 0.1} | Random Search      'model\_\_filters1': [32, 64, 128],      'model\_\_filters2': [64, 128, 256],      'model\_\_kernel\_size1': [3, 5],      'model\_\_kernel\_size2': [3, 5],      'model\_\_dropout\_rate': [0.1, 0.2, 0.3] |
| AutoEncoder with Best deep learning Model: BIGRU | RMSE: 3.7464 | MAE: 3.3055 | MAD: 3.3055 | MAPE: 15.7060% | Lag : 24 |  |
| BIGRU with feature extraction | Test RMSE: 3.6559 |  |  |  | Lag : 24 |  |
| Ensamble of BIGRU and BILSTM | RMSE: 3.6920 | MAE: 3.0036 | MAD: 3.0036 | MAPE: 15.3549% | Lag : 24 |  |
| BIGRU with parameters of optimization Bayesian | RMSE: 3.8353 | MAE: 3.4039 | MAD: 3.4028 | MAPE: 15.9088% | Best units first layer: 288  Best dropout first layer: 0.0  Best units second layer: 128  Best dropout second layer: 0.1  Best learning rate: 0.00048467141838873354 | Bayesian Optimization |

**Question 1:** I ran linear regression, Polynomial Regression with PCA (I used to reduce dimensions and improve performance), Moving average for Y and Moving Average with all variables, Arima for and Arima with all variables, and Gaussian process with PCA. In addition, in the plot of ACF and PACF, there is no significant seasonality.

**Question 2:** I run KNN(n\_neighbors = [3, 5, 7, 9, 11, 13, 15, 17]), Decision tree (max\_depths = [3, 5, 7, 9, 11, 13, 15, 17]) and SVR .

**Question 3:** I run GRU, LSTM, BILSTM, BIGRU, and CNN with lag 24.

**Question 4:** To tune models, I used a random search optimization. Random Search is used to efficiently explore a broad hyperparameter space for various deep learning models like GRU, LSTM, BiLSTM, BiGRU, and CNN and other machine learning models, offering a good balance between computational cost and parameter optimization.

In addition, to improve the best model of deep learning I used Bayesian Optimization.Bayesian Optimization then refines these models by targeting promising hyperparameter regions, enhancing model performance with fewer evaluations and higher precision. (All parameters are mentioned in the table)

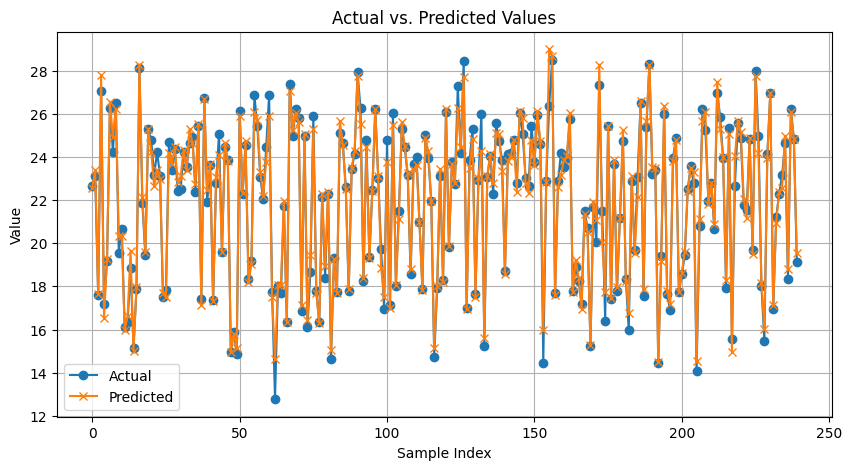
**Question 5:**

* In some models I used PCA to reduce dimensions and improve them. In some models, I used stacking, stacking with PCA, AutoEncoder, and optimization Bayesian.
* Statistical model: I run Stacking- With best ST models including Linear regression, Polynomial Regression, Gaussian process with PCA, and Arima with best parameters, and Stacking- With best ST models - PCA. And, in the end, Stacking- With the best ST models-PCA is the best.
* Machine learning: I run Stacking-Best ML models and Stacking-Best ML models-PCA. And the best model is Stacking-Best ML models -PCA. (KNN, Decision tree, and SVR which are optimized with random search)
* Deep learning: I run AutoEncoder with Best deep learning Model-BIGRU, BIGRU with feature extraction, Ensamble of BIGRU and BILSTM, BIGRU with parameters of optimization Bayesian. But, the best model is BIGRU-Optimized with random search.

1. PCA (Principal Component Analysis): Used to reduce the dimensionality of the dataset, improving model training times and helping to prevent overfitting by removing less informative variables. PCA simplifies the complexity in high-dimensional data while retaining trends and patterns.
2. Stacking: This ensemble method combines multiple classification or regression models via a meta-classifier or meta-regressor. Stacking is used to increase the prediction accuracy by taking into account diverse perspectives from various models, leading to better generalization on unseen data.
3. Stacking with PCA: Integrating PCA into stacking reduces the input feature space, which can lead to more efficient learning and less overfitting in the stacked models. It makes the stacking ensemble not only faster but potentially more accurate by focusing on the most significant features.
4. AutoEncoder: Primarily used for unsupervised learning of efficient codings, AutoEncoders are helpful in dimensionality reduction and feature learning, especially useful in preparing data for complex predictive models by highlighting intrinsic structures or patterns in the data.
5. Bayesian Optimization: Employed for optimizing hyperparameters, Bayesian optimization uses a probabilistic model to guide the search for the best hyperparameters. This technique is particularly effective because it balances the exploration of the parameter space with the exploitation of known good regions, making it superior for finding optimal solutions efficiently.

**Question 6:**

* Statistical model: I run linear regression, Polynomial Regression with PCA (I used to reduce dimensions and improve performance), Moving average for Y and Moving Average with all variables, Arima for and Arima with all variables, and Gaussian process with PCA. After random search optimization, the best Model is Arima Multivariate (3, 2, 1) and linear regression.



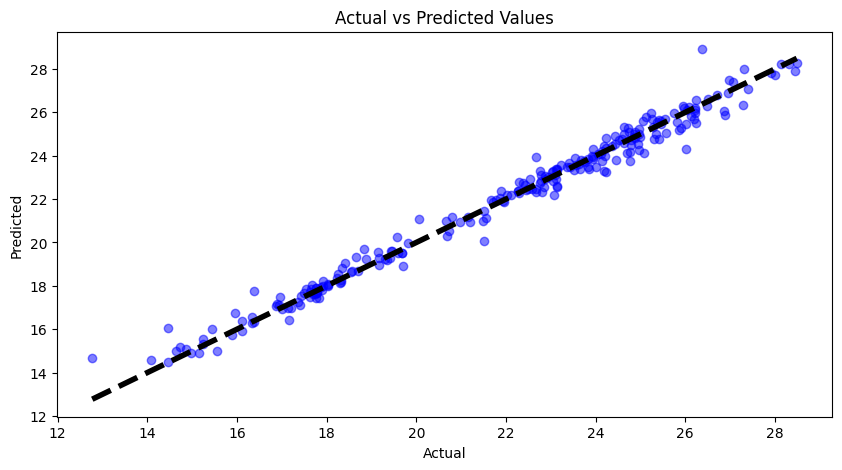
Because, Metrices including RMSE, MAE, MAD, and MAPE are 0.474, 0.321, 0.321, 1.502 for Arima, and 0.47, 0.32, 0.32, 1.52 respectively, and the lowest. To improve statistical models, I run Stacking- With best ST models including Linear regression, Polynomial Regression, Gaussian process with PCA, and Arima with best parameters, and Stacking- With best ST models - PCA. Stacking- With best ST models-PCA is the best. The best model has metrics:

RMSE: 0.46

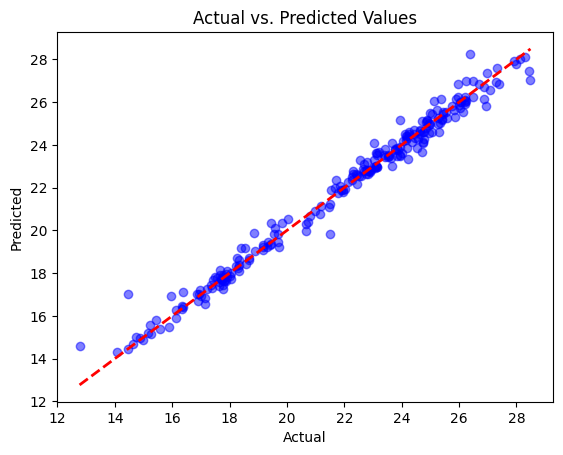
MAE: 0.32

MAD: 0.21

And the lowest.



* Machine learning: I run KNN(n\_neighbors = [3, 5, 7, 9, 11, 13, 15, 17]), Decision tree (max\_depths = [3, 5, 7, 9, 11, 13, 15, 17]) and SVR with PCA. After random search optimization, the best model is KNN -Optimized with {'n\_neighbors': 2, 'p': 2, 'weights': 'distance'}with metrics RMSE, MAE, MAD: 0.45, 0.30, 0.21 respectively, which is the lowest.



Then, to improve it, Stacking-Best ML models (KNN, Decision tree, and SVR which are optimized with random search) , Ensemble with Random Forrest, Stacking-Best ML models-PCA. Stacking-Best ML models-PCA is the best model. The metrics of the best model:

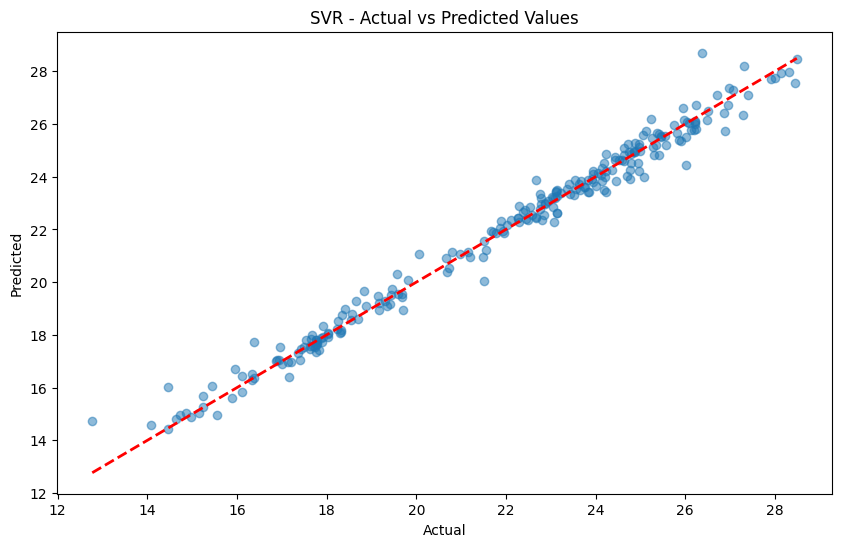
RMSE: 0.4137

MAE: 0.2824

MAD: 0.2067

MAPE : -1.3394%

Which is the lowest.



* Deep learning: I run GRU, LSTM, BILSTM, BIGRU, and CNN with lag 24. After random search optimization, the best model BIGRU -Optimized has RMSE, MAE, MAD, MAPE : 3.5980, 3.0885, 3.1159, 15.3449 respectively which is the lowest. Then to improve the best model, I tried AutoEncoder with the Best deep learning Model- BIGRU and BIGRU with parameters of optimization Bayesian. But, the best model is BIGRU -Optimized with random search.

RMSE: 3.5980

MAE: 3.0885

MAD: 3.1159

MAPE: 15.3449

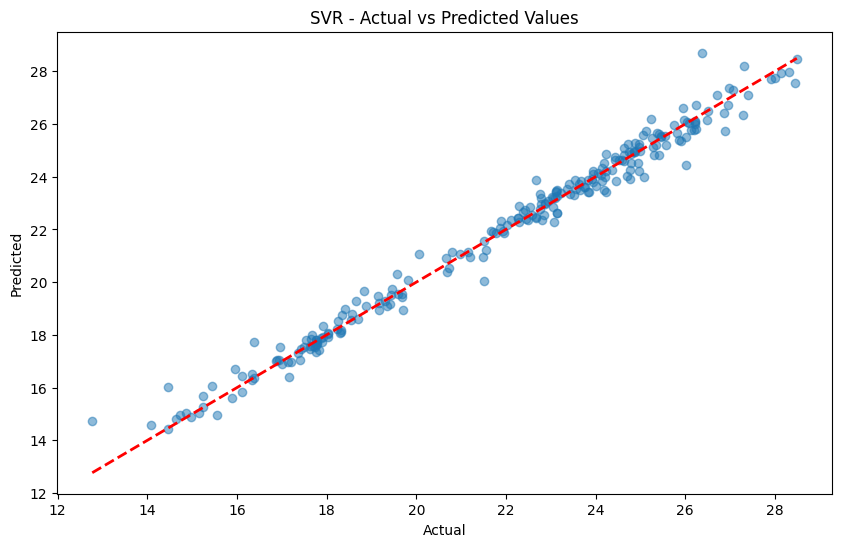
All Models: the best model is The Stacking-Best ML models-PCA is the best model (KNN, Decision tree and SVR which are optimized with random search) with performance:

RMSE: 0.4137

MAE: 0.2824

MAD: 0.2067

MAPE : -1.3394%



**Question 7:**

* Statistical model: After random search, the worst models including Arima and Moving Average (except Arima-Multi) are not good because they have the highest metrics.
* Machine learning: After random search optimization, Decision Tree{'max\_depth': 5, 'min\_samples\_leaf': 9, 'min\_samples\_split': 6} has the highest metrics including RMSE, MAE, MAD, MAPE: 0.49, 0.34, 0.24, 1.60 respectively.
* Deep learning: After random search process, the worst model is CNN {'model\_\_kernel\_size2': 3, 'model\_\_kernel\_size1': 5, 'model\_\_filters2': 256, 'model\_\_filters1': 32, 'model\_\_dropout\_rate': 0.1}with the highest metrics including RMSE, MAE, MAD, and MAPE: 3.9690, 3.2711, 3.2745, 16.2689 respectively.

All Models: the worst model is CNN with the highest metrics including RMSE, MAE, MAD, and MAPE: 3.9690, 3.2711, 3.2745, 16.2689 respectively.