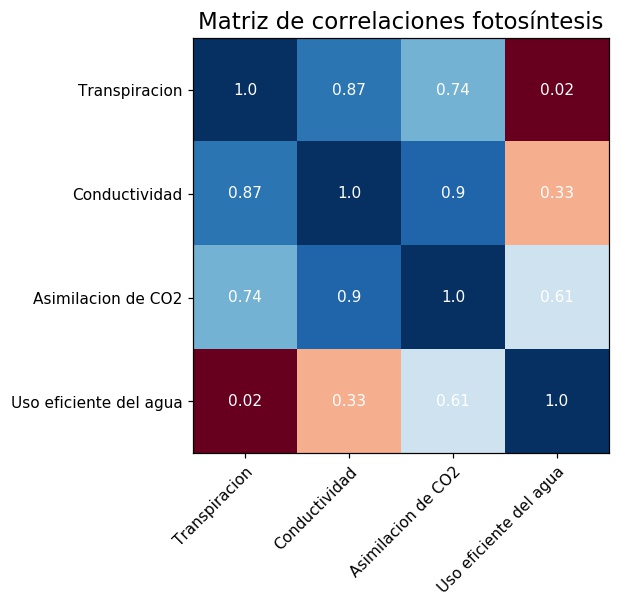
**SELECCIÓN VARIABLES UTILIZADAS**

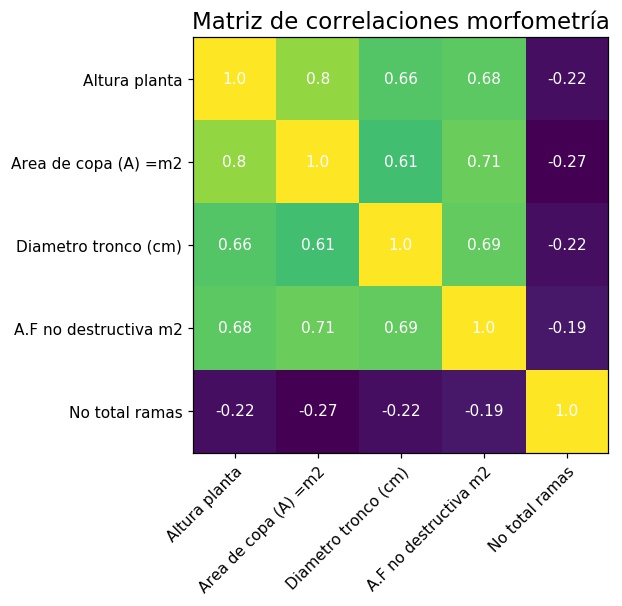
Para la selección de variables, se realizó una correlación para cada conjunto de datos con el fin de descartar las variables con mayor similitud entre ellas (mayor coeficiente de correlación) y así poder analizar para cada modelo aquellas variables con menor coeficiente de correlación, es decir, aquellas con menor similitud entre ellas.

A continuación, se muestran las matrices de correlación y las variables seleccionadas para cada conjunto de datos

1. **Correlación fotosíntesis**



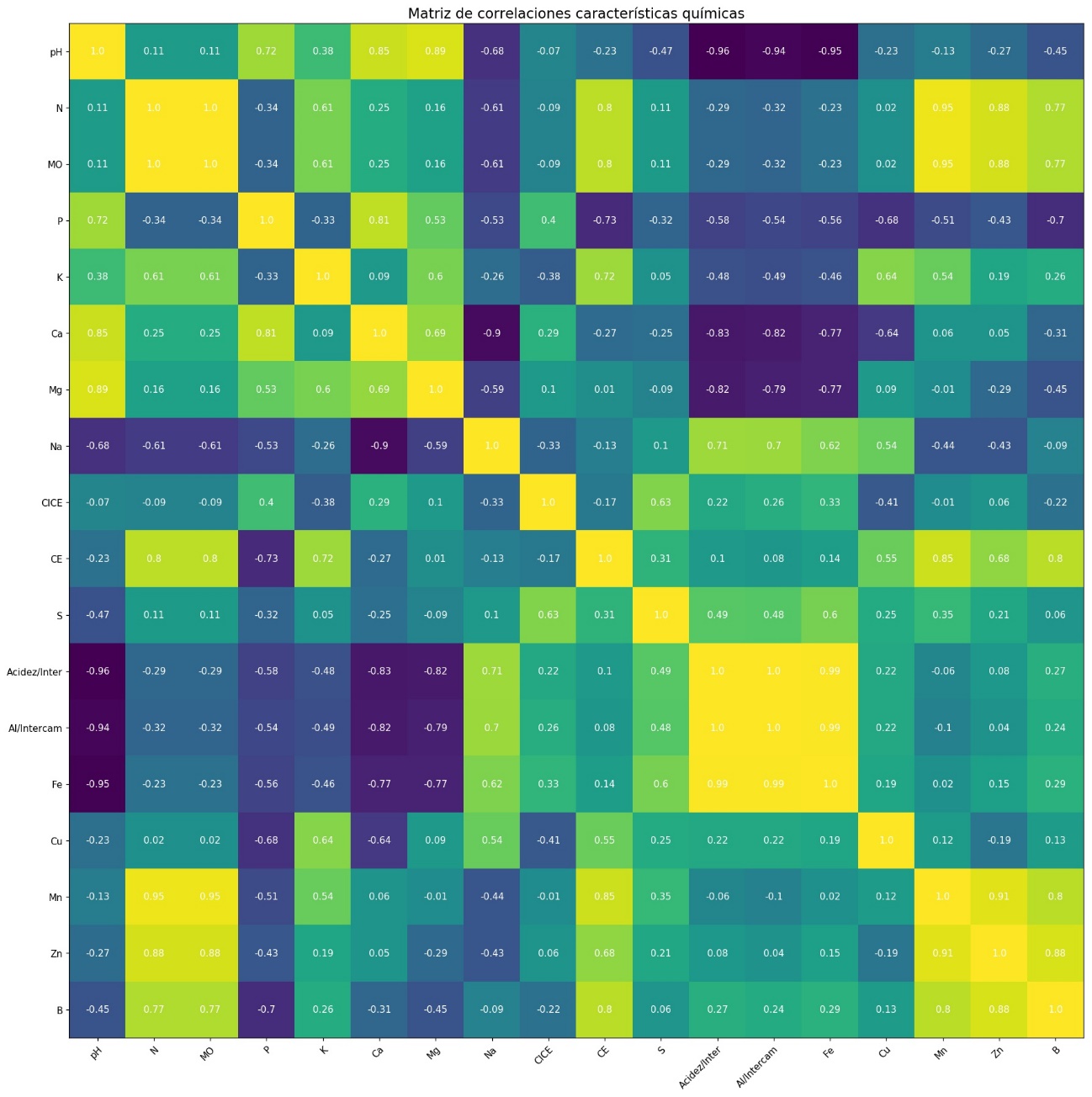
En base a la correlación del conjunto de datos de las variables de fotosíntesis, se toman en cuenta las variables **“transpiración” y “uso eficiente de agua”,** teniendo en cuenta que estas son las que presentan menor correlación y por lo tanto menor similitud entre ellas.

1. **Correlación morfometría**

En base a la correlación del conjunto de datos de las variables de morfometría, se toman en cuenta las variables “**Diámetro del tronco”** teniendo en cuenta que estas son las que presentan menor correlación y por lo tanto menor similitud entre ellas.

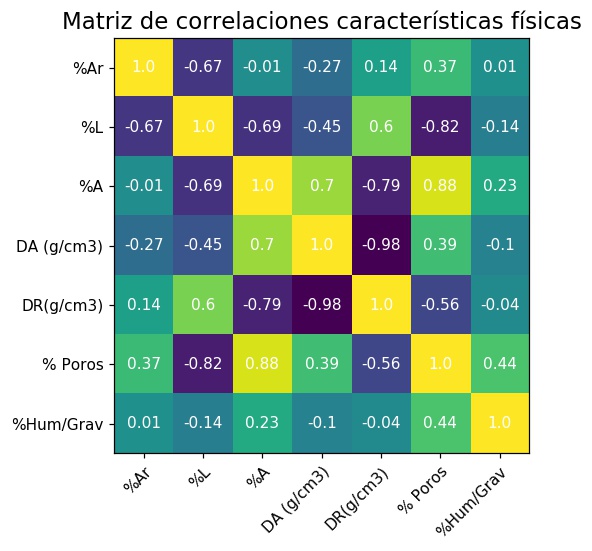
**Nota:** Quitamos NÚMERO TOTAL DE RAMAS

1. **Correlación características químicas del suelo**



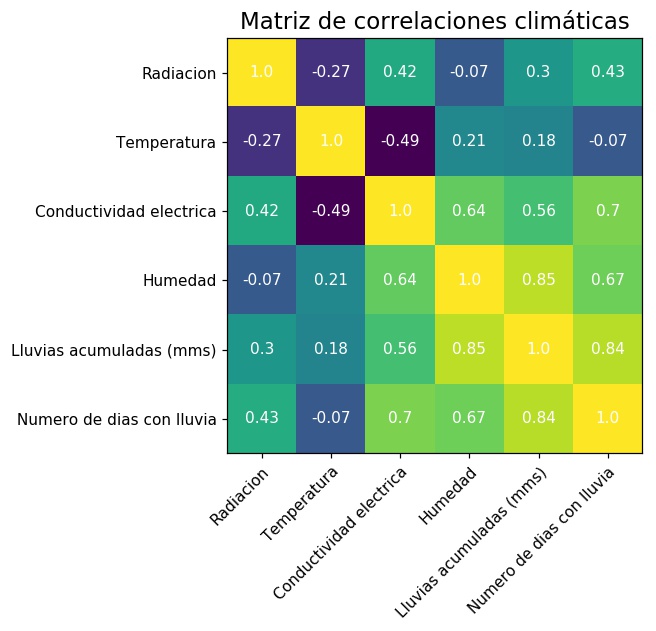
En base a la correlación del conjunto de datos de las características químicas, se toman en cuenta las variables **P** (en vez de pH, Ca), **MO** (IGUAL a N; en vez de K, CE), **Mg** (En vez de Ca y K) **Na** (en vez de Al/Intercam, Acidez, Fe y Cu); Respecto a cada una de ellas, miramos qué variables tenían un coeficiente de correlación grande para así descartar las variables que lo tuvieran del modelo.

1. **Características físicas del suelo**



En base a la correlación del conjunto de datos de las características físicas, se toman en cuenta las variables **“%A”, “%Hum/Grav”** teniendo en cuenta que estas son las que presentan menor correlación y por lo tanto menor similitud entre ellas.

1. **Climáticas**



En base a la correlación del conjunto de datos de las climáticas, se toman en cuenta las variables “**Radiación”, “Temperatura” y “humedad”** teniendo en cuenta que estas son las que presentan menor correlación y por lo tanto menor similitud entre ellas.

**Nota:** A pesar de la alta correlación de la variable “Lluvias acumuladas” respecto de las demás, se incluyó en el modelo debido a la opinión por parte de los expertos; los cuales indican que es una variable a analizar debido a la presunta incidencia que se observa experimentalmente en los cultivos.

**MODELO LINEAL GENERALIZADO**

Para el modelo lineal generalizado se trabaja con las variables independientes seleccionadas, mencionadas anteriormente. En primera instancia, el modelo GLM se corrió con todas ellas, como se muestra a continuación.

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | Rendimiento | **No. Observations:** | 390 |
| **Model:** | GLM | **Df Residuals:** | 377 |
| **Model Family:** | Gamma | **Df Model:** | 12 |
| **Link Function:** | log | **Scale:** | 0.65833 |
| **Method:** | IRLS | **Log-Likelihood:** | -3308.5 |
| **Date:** | Wed, 28 Nov 2018 | **Deviance:** | 266.94 |
| **Time:** | 14:26:46 | **Pearson chi2:** | 248. |
| **No. Iterations:** | 15 | **Covariance Type:** | nonrobust |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **z** | **P>|z|** | **[0.025** | **0.975]** |
| **Uso eficiente del agua** | 0.0190 | 0.049 | 0.384 | 0.701 | -0.078 | 0.116 |
| **Transpiracion** | 0.1403 | 0.072 | 1.954 | 0.051 | -0.000 | 0.281 |
| **Diametro tronco (cm)** | 0.0607 | 0.015 | 3.998 | 0.000 | 0.031 | 0.091 |
| **P** | 0.9526 | 0.151 | 6.295 | 0.000 | 0.656 | 1.249 |
| **MO** | -0.5472 | 0.473 | -1.157 | 0.247 | -1.474 | 0.379 |
| **Na** | -2.1315 | 7.464 | -0.286 | 0.775 | -16.761 | 12.498 |
| **Mg** | 2.9475 | 0.407 | 7.245 | 0.000 | 2.150 | 3.745 |
| **%A** | -0.2185 | 0.046 | -4.777 | 0.000 | -0.308 | -0.129 |
| **%Hum/Grav** | 0.4094 | 0.084 | 4.847 | 0.000 | 0.244 | 0.575 |
| **Radiacion** | 6.091e-07 | 1.09e-07 | 5.587 | 0.000 | 3.95e-07 | 8.23e-07 |
| **Temperatura** | -0.2819 | 0.075 | -3.774 | 0.000 | -0.428 | -0.135 |
| **Humedad** | -35.3668 | 4.657 | -7.595 | 0.000 | -44.494 | -26.240 |
| **Lluvias acumuladas (mms)** | 0.0090 | 0.001 | 6.983 | 0.000 | 0.006 | 0.012 |

Cada una de las variables tiene un valor p; el cual indica que las que tengan valores por debajo de 0.05, son aquellas variables que tienen mayor incidencia en la variable respuesta, las cuales son:

**'Diámetro tronco (cm)', 'P', 'Mg', '%A', '%Hum/Grav', 'Radiación', 'Temperatura', 'Humedad', 'Lluvias acumuladas (mms)'**

Es por esto último que posteriormente se entrena y valida el modelo con las variables de valores p menores que 0.05; aquellas que mayor incidencia tienen en la variable respuesta. A continuación, se muestran los resultados

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | Rendimiento | **No. Observations:** | 312 |
| **Model:** | GLM | **Df Residuals:** | 303 |
| **Model Family:** | Gamma | **Df Model:** | 8 |
| **Link Function:** | log | **Scale:** | 0.68987 |
| **Method:** | IRLS | **Log-Likelihood:** | -2650.2 |
| **Date:** | Wed, 28 Nov 2018 | **Deviance:** | 226.34 |
| **Time:** | 14:26:51 | **Pearson chi2:** | 209. |
| **No. Iterations:** | 100 | **Covariance Type:** | nonrobust |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **z** | **P>|z|** | **[0.025** | **0.975]** |
| **Diametro tronco (cm)** | 0.0622 | 0.016 | 3.877 | 0.000 | 0.031 | 0.094 |
| **P** | 1.0123 | 0.166 | 6.084 | 0.000 | 0.686 | 1.338 |
| **Mg** | 2.5123 | 0.392 | 6.412 | 0.000 | 1.744 | 3.280 |
| **%A** | -0.2323 | 0.046 | -5.078 | 0.000 | -0.322 | -0.143 |
| **%Hum/Grav** | 0.3997 | 0.083 | 4.787 | 0.000 | 0.236 | 0.563 |
| **Radiacion** | 4.662e-07 | 8.58e-08 | 5.431 | 0.000 | 2.98e-07 | 6.34e-07 |
| **Temperatura** | -0.2630 | 0.058 | -4.551 | 0.000 | -0.376 | -0.150 |
| **Humedad** | -34.5760 | 4.985 | -6.936 | 0.000 | -44.347 | -24.805 |
| **Lluvias acumuladas (mms)** | 0.0083 | 0.001 | 6.622 | 0.000 | 0.006 | 0.011 |

Como se puede observar en la tabla previamente presentada, de los valores de p menores a 0.05, las variables: **'Diámetro tronco (cm)', 'P', 'Mg', '%A', '%Hum/Grav', 'Radiación', 'Temperatura', 'Humedad', 'Lluvias acumuladas (mms)'** son aquellas variables representativas dentro del modelo.

De cada uno de sus coeficientes se concluye que:

Para la variable **'Diámetro tronco (cm)',** con un coeficiente positivo de 0.0622, se deduce que parece influir positivamente en la variable dependiente rendimiento.

La variable ‘**P',** con un coeficiente positivo de 1.0123, parece influir positivamente en el rendimiento.

La variable **‘Mg',** con un coeficiente positivo de 2.5123, parece influir positivamente en la variable dependiente.

La variable **‘%A',** con un coeficiente negativo de -0.2323, parece influir negativamente en el rendimiento del cultivo.

De los coeficientes positivos de las variables ‘**%Hum/Grav', 'Radiación' y 'Lluvias acumuladas,** de 0.3997, 4.662e-07 y 0.0083 respectivamente, se puede deducir que estrían influyendo positivamente en la variable dependiente.

De las variables **'Temperatura' y 'Humedad'** con coeficientes de -0.2630 y -34.5760 respectivamente, se podría estar concluyendo que estas variables en elevada proporción, estarían influyendo negativamente al rendimiento del cultivo.

Posterior al entrenamiento, se realizó una validación de los modelos con las metodologías Cross Validation y Hold Out…

**MÁQUINAS DE SOPORTE VECTORIAL**

Para efectos de comparación entre los dos modelos, el modelo de máquinas de soporte vectorial se entrenó y validó con las mismas variables independientes del modelo GLM; aquellas que mostraron mayor incidencia en la variable dependiente.

(luego de escalar las variables y encontrar los mejores parámetros para las máquinas de regresión...) el modelo se entrena y valida con un kernel lineal; generando para él las predicciones para el rendimiento.

Los coeficientes generados en las máquinas de soporte vectorial para las variables **'Diámetro tronco (cm)', 'P', 'Mg', '%A', '%Hum/Grav', 'Radiación', 'Temperatura', 'Humedad', 'Lluvias acumuladas (mms)',** son 460.47348037, 1676.49904291, 739.47395887, -1014.12586028, 885.0244371, 538.83755418, -278.53260164, -1549.99072426, 1468.66102623 respectivamente

*Cabe aclarar que del signo de los coeficientes para cada una de las variables del modelo SVM, se observa que la incidencia comparada con el modelo anterior es similar.*

Luego de generar los coeficientes para cada variable se concluye que:

Para la variable **'Diámetro tronco (cm)',** con un coeficiente positivo de 460.47348037, se deduce que parece influir positivamente en la variable dependiente rendimiento.

La variable ‘**P',** con un coeficiente positivo de 1676.49904291, parece influir positivamente en el rendimiento.

La variable **‘Mg',** con un coeficiente positivo de 739.47395887, parece influir positivamente en la variable dependiente.

La variable **‘%A',** con un coeficiente negativo de -1014.12586028, parece influir negativamente en el rendimiento del cultivo.

De los coeficientes positivos de las variables ‘**%Hum/Grav', 'Radiación' y 'Lluvias acumuladas,** de 885.0244371, 538.83755418 y 1468.66102623 respectivamente, se puede deducir que estrían influyendo positivamente en la variable dependiente.

De las variables **'Temperatura' y 'Humedad'** con coeficientes de -278.53260164, -1549.99072426 respectivamente, se podría estar concluyendo que estas variables en elevada proporción, estarían influyendo negativamente al rendimiento del cultivo.