**\section{Γραμμική παρεμβολή και παλινδρόμηση}**

Αρχικά ελέγχουμε την σχέση \tl{SMP-ngas} με ένα διάγραμμα και δημιουργούμε το \tl{trendline} (το οποίο ενσωματώσαμε στο διάγραμμα). Ύστερα από την σχέση αυτή υπολογίζουμε τα αντίστοιχα σφάλματα.

\begin{figure}

\makebox[\textwidth][c]{**\includegraphics**[width=1.4\textwidth]{fig/smp-ngas.png}}%

\caption{\tl{SMP and ngas}}

**\label{figure:15}**

\end{figure}

\begin{minted}[linenos]{python}

smp = train\_imp[:,0]

ngas = dataset[:,6]

plt.scatter(ngas,smp, label = 'smp and ngas values')

z = np.polyfit(ngas, smp, 1)

p = np.poly1d(z)

plt.plot(ngas,p(ngas),'r', label='smp = 0.000802\*ngas + 25.160520')

plt.ylabel('smp')

plt.xlabel('ngas production')

plt.title('smp for different quantities of lignite and ngas')

plt.legend(loc='upper left')

plt.show()

print 'smp = %.6f\*ngas + %.6f'%(z[0],z[1])

\end{minted}

Παρατηρούμε πως μόνο από γραμμική παρεμβολή μεταξύ των δύο στοιχείων έχουμε ποσοστό 30\% για πρόβλεψη της τιμής με μέσο απόλυτο σφάλμα τα 9.6\euro. \\

\tl{Explained variance score = 0.323114 \\

Mean absolute error = 9.655872 \\

Mean squared error = 154.212626 \\

R2\\_score = 0.323114} \\

Κάνοντας την ίδια διαδικασία με άλλες μεταβλητές παρατηρούμε τιμές κοντά ή χαμηλότερες από το \tl{importance score} του \tl{random forest}. Παρατηρούμε επίσης πως καλύτερα σκορ παρουσιάζουν ξανά οι μεταβλητές \tl{waters} και \tl{load\\_forecast}. Να σημειωθεί εδώ πως το μοντέλο αυτό κάνει απλά γραμμική παρεμβολή των δεδομένων σε σχέση με το \tl{SMP} και σε καμία περίπτωση δεν μπορεί να χρησιμοποιηθεί για πρόβλεψη και έγινε μόνο για καλύτερο έλεγχο των χαρακτηριστικών. Είναι όμως ενδιαφέρον το γεγονός πως οι μεταβλητές παρουσιάζουν παρόμοια συμπεριφορά ως προς το πόσο καλά μπορούν να καθορίσουν την έξοδο με τον αλγόριθμο για \tl{random forest}.

\begin{table}[h!]

\centering

\begin{tabular}{||c c c c||}

\hline

\tl{Feature} & \tl{Variance} & \tl{MAE} & \tl{MSE}\\

\hline\hline

\textlatin{availability} & 0.007312 & 11.755722 & 226.160655 \\

\hline

\textlatin{exports} & 0.009829 & 11.669390 & 225.587189\\

\hline

\textlatin{hydrogen} & 0.042356 & 11.451819 & 218.176712 \\

\hline

\textlatin{imports} & 0.070360 & 11.254051 & 211.796578 \\

\hline

\textlatin{lignite} & 0.038649 & 11.459376 & 219.022228 \\

\hline

\textlatin{load\\_forecast} & 0.134780 & 10.836657 & 197.120519 \\

\hline

\textlatin{ngas} & 0.323114 & 9.655872 & 154.212626 \\

\hline

\textlatin{res\\_forecast} & 0.059719 & 11.166531 & 214.221031\\

\hline

\textlatin{waters} & 0.142597 & 10.983204 & 195.339044 \\

\hline

\textlatin{waip} & 0.069096 & 11.345222 & 212.084584 \\

\hline

\end{tabular}

**\label{table:3}**

\caption{Γραμμική παρεμβολή των χαρακτηριστικών ως προς το \tl{SMP}}

\end{table}