

# PREDIÇÃO DO TEOR DE SILÍCIO NO FERRO-GUSA UTILIZANDO TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL

Bernardo Penna Resende de Carvalho<sup>(1)</sup>

André Luiz de Senna<sup>(2)</sup>

Fábio Mayrink Morais<sup>(3)</sup>

## RESUMO:

Apresentamos neste trabalho um sistema baseado em técnicas de Inteligência Artificial (IA) combinadas de maneira inédita na literatura para fazer a predição do teor de silício em amostras de gusa retiradas de corridas do alto-forno 2 da Acesita. O teor de silício é um dos principais indicadores de qualidade do gusa, portanto um sistema capaz de fazer predições acuradas permite tomar ações corretivas mais rapidamente melhorando a qualidade do processo.

O sistema utiliza uma combinação de machine learning, redes neuronais e algoritmo de otimização bayesiano e permite fazer a predição online do teor de silício no gusa da próxima corrida a partir da interpretação das informações sobre forno, processo e análises mineralógicas e químicas das matérias-primas. Os modelos são construídos exclusivamente com base em informações históricas do banco de dados do supervisório e dos laboratórios.

São utilizadas ferramentas para seleção de dados relevantes, sem intervenção do usuário, que eliminam ruído e redundância dos dados do supervisório. Tais ferramentas exploram características variadas dos dados, e quando usadas em conjunto, geram um conjunto de dados consistente para utilização nos modelos de I.A.

**PALAVRAS-CHAVE:** Inteligência artificial, predição, qualidade

---

<sup>(1)</sup> Engenheiro Eletricista - UFMG, Mestrando em Engenharia Elétrica – UFMG, Analista de Pesquisa e Desenvolvimento - Divisão de I.A. da Vetta Technologies.

<sup>(2)</sup> Graduado e mestre em Ciência da Computação – UFMG, PhC em Ciência da Computação - UFMG, Diretor Técnico - Vetta Technologies.

<sup>(3)</sup> Bacharel em Física - UFMG, MSc Engenharia Metalurgica e de Minas – UFMG, Gerente Técnico de Redução – Acesita S.A..

# 1 Introduction

In this work we present a system based on a novel combination of Artificial Intelligence (A.I.) techniques to predict the silicon grade of samples of pig iron from blast furnace 2 runners in Acesita. Silicon grade is one of the main indicators of pig iron quality, and therefore a system capable of making accurate predictions allows the furnace manager to take corrective action more quickly, thus improving the process quality.

Changes in operational conditions of the blast furnace may interfere in the grade of silicon in an unpredictable way. The furnace manager may take several hours to succeed turning the value back to normal ranges. Such a delay may cause one or two runs to have pig iron with silicon grade out of the expected specification. We built a decision support system that allow the manager to forecast the silicon grade in the next run, reducing the variability of this indicator in the final product therefore, improving the quality of the production process.

The system was built using the Vetta Artificial Intelligence Framework (VAIF), which uses a combination of several artificial intelligence techniques (neural networks, bayesian optimization algorithm, genetic programming among others) integrated to allow maximum effectiveness in data analysis. Data were taken *in natura* directly from the supervisory system's database without any type of mathematical or statistical treatment, covering a period of abnormal blast furnace functioning.

## 2 Vetta Artificial Intelligence Framework

The VAIF is a software system developed at Vetta Technologies that implements several Artificial Intelligence (A.I.) techniques in the same architecture. VAIF can be used to perform data analysis using the best of each A.I. technique to find patterns and valuable information in massive databases. The integration of different algorithms cooperating to perform the same task allows the system to reach results that are more accurate and reliable than the ones obtained by using any of the techniques separately.

The framework can be applied in many different domains to analyze any kind of data. In addition to the A.I. algorithms that actually process the information, it has filtering algorithms to deal with noise, dimensionality reduction algorithms to allow the use of very large amounts of data, statistical operators to perform pre-analysis simplification and several other mathematical devices to treat specific or anomalous characteristics of the database.

This gamut of algorithmic tools makes VAIF a powerful and versatile framework that can be tuned to solve optimization, forecasting, pattern extraction or any other general data analysis problem issued at industrial plants. It is powerful because it reaches accurate results for any class of problem. It is versatile because it can be applied over raw data collected from the supervisory system (or any support system) without need for pre-processing or filtering.

### **3 Modeling the problem**

We made the prediction models combining A.I. algorithms with dimensionality reduction methods. To validate results, we used a database generated by the Acesita's supervisory system between February-2002 and July-2002 for blast furnace 2. Data were used *in natura*, i.e., as they were stored in the database. This period was a period of abnormal functioning of the blast furnace. We chose it because it is harder to build prediction models for this period than for periods of normal functioning, so it is more suitable for use in a proof of concept.

The database is a set of tables with a total of 803 columns. Each of them is a variable sampled a given number of times (lines in the table). One of these variables is the silicon grade in the pig iron.

We used a kernel method [10] integrated with a bayesian optimization algorithm [7] to perform predictions. Kernel methods are an evolution of Artificial Neural Networks (ANNs) [8] that usually achieves more accurate and reliable results. Like ANNs, Kernel Methods are very sensitive to parameter tuning and to the overall quality of training data, so it is important to integrate it with an optimization algorithm that takes care of the process of finding optimal parameters and with a data pre-processor to assure that training data are relevant and non-noisy.

#### **3.1 Kernel Methods**

Kernel Methods (KM) are learning machines similar to multi-layer neural networks with only one hidden layer that provides a non-linear mapping of data at some feature space. This mapping is done implicitly using kernel functions with different characteristics. It uses a supervised learning process that guarantees unique solutions to each problem.

KM force the construction of an optimal decision hyper-plane by the usage of structural risk minimization theory [9], which leads to a good performance at unseen data. It has a special feature that is the automatic adjustment of parameters in order to match to the problem complexity. KM are capable of recognizing the most important set of data of a problem, the so-called support vectors, which are the only set of data that effectively are needed to build the decision hyper-plane.

#### **3.2 Bayesian Optimization Algorithm**

Bayesian Optimization Algorithm (BOA) [7] is an A.I. searching technique that applies principles of population evolution and natural selection to evolve sets of solutions for optimization problems. The algorithm starts with an initial population of possible solutions generated either randomly or by some particular heuristics and evolve this population by combining its individuals using approximations of Bayesian Networks.

BOA is an evolution of Genetic Algorithms [2] and applies the principles of genetics and of Darwin's evolution of species theory. However, BOA usually achieves better results (at larger computational cost) and can be applied with success for a number of problems where regular Genetic Algorithms fail.

### 3.3 Dimensionality reduction methods

The database generated by the supervisory system contains a very large number of variables sampled several times per hour (some are sampled once per second). Therefore the system needs to perform some kind of pre-processing to allow the A.I. predictor to deal with such a large amount of data.

We applied dimensionality reduction techniques to obtain a sub-set of the original set of variables that satisfies the trade-off between the size and the quality of the data presented to the A.I. predictor. It is desired that the sub-set of data contain the maximal of relevant information from the original set without useless redundancy or noise.

We performed the experiments using the following approaches.

1. *Ad hoc* – variables selected by hand using domain knowledge.
2. SPCA – Simple Principal Component Analysis [6], a decomposition technique that projects the original points in an induced space defined by eigenvectors of the original representation. The difference between SPCA and regular PCA is that the former uses hebbian learning [5] to approximate the eigenvectors.
3. RFE – Recursive Feature Elimination [3], a feature selection technique that uses a Kernel Method (KM) to rank and classify each variable. The method uses the weights obtained in a training round of a KM as comparison criterion to recursively remove unimportant variables.
4. Selection by Pearson Correlation [4] – a feature selection method that uses Pearson's correlation to select variables that have small correlation among them and are highly correlated to the output.

A complete survey of dimensionality reduction techniques can be found in [1].

## 4 Results

We divided the database in two groups: the training set and the testing set. The former is composed of 80% of the data, while the remaining 20% constitutes the latter. The training set is used to build the prediction model while the testing set is used to validate results. This distinction is important because the accuracy of the prediction model can only be properly measured if we evaluate it using data that was never presented to the model before.

Variations in the input variables affect the silicon grade with different time delays, i.e., some of them may take a few minutes to influence the output while others may take a few hours to do it. We incorporate this physical behavior in the system by introducing a time shift in some variables, i.e., the value  $V_t$  for a given variable  $V$  at time  $t$  is used as input of the system only at time  $t+\Delta$ . Different variables have different  $\Delta$ 's. The system does not assume a fixed value for each  $\Delta$ , but uses the A.I. algorithms to select them properly as part of the solution for the problem. It is a simple operation that synchronizes all the inputs according to the expected delay of each input to affect the output variable.

We performed six experiments considering different choices of sets of variables.

1. Full database – all the variables
2. Ad hoc choice of data (A) – sets of variables were selected using domain knowledge.
3. Ad hoc choice of data (B) – the input of the prediction model is the output variable itself shifted a given number of minutes in time.
4. SPCA
5. RFE
6. Pearson Correlation

All experiments were performed twenty times, and the presented results are always their mean values. In the tables below, “Data ratio”, is the percentage of variables used at the experiments. “Average error” is calculated by comparing expected values against the model’s output for the test set averaging the relative errors over all examples:

$$\frac{\sum_n \left( \frac{|Expected_i - Output_i|}{Expected_i} \right)}{n}$$

## Experiment 1 - Full database

The first results are obtained when using all available data without pre-processing.

Table 1: Full database

Time Shift (hh:mm)	Average error (%)
00:30	40.45
01:00	40.55
02:00	40.43
06:00	40.48

The results show the importance of an adequate dimensionality reduction analysis to reduce the number of used variables and minimize the average error. The results in Table 1 with a high average error regardless of the time shifts show that predictions using the full database are meaningless. This happens because there is a lot of redundancy and noise in the database, which makes harder the task of building good predictors.

## Experiment 2 - Ad hoc choice of data (A)

As a first attempt to obtain results for a reduced data set, the variables were chosen using some *a priori* knowledge about the process. We selected some groups of variables and performed the experiments.

Table 2: Ad hoc choice of data (A)

Data sets	Data ratio (%)	Average error (%)
Chemical analysis	08.34	17.02
Grain sized analysis	13.82	18.23
Chemical and grain sized analysis	20.55	18.82
Refrigeration sensors	21.42	40.74
Supervisory data set 1	40.60	27.77
Supervisory data set 2	12.20	37.66
Supervisory data set 3	81.20	40.43

Laboratory variables, mainly the chemical analysis, have significant importance when compared to others, as seen in Table 2. But it does not mean that variables in the other sets are not relevant to make the prediction.

### Experiment 3 - Ad hoc choice of data (B)

We experimented using only the silicon grade at different moments in the past to forecast future values.

Table 3: Ad hoc choice of data (B)

Time shift (hh:mm)	Average error (%)
00:30	16.77
01:00	15.86
02:00	17.23
06:00	24.00

Table 3 reveals that the output feedback at different time shifts is very important to achieve good accuracy. Actually, this variable alone is enough to make good predictors.

### Experiment 4 - SPCA

The SPCA method was performed using different sets of parameters because it is known that its performance is very sensitive to parameter tuning.

Table 4: SPCA

Case	Data ratio (%)	Average error (%)
Parameters set 1	4.36	29.18
Parameters set 2	4.36	19.61
Parameters set 3	4.36	19.90
Parameters set 4	4.36	19.16

### Experiment 5 - RFE

The tests of the RFE method are divided in two groups, based on the existence of a pre-processing data phase, which is a linear normalization, where the range of data is rearranged to transform its minimum value to 0 and its maximum value to 1.

In the following results, data normalization is not implemented.

Table 5: RFE without normalization

Time shift (hh:mm)	Data ratio (%)	Average error (%)
00:30	4.98	40.45
01:00	4.98	40.43
02:00	4.98	40.45
06:00	4.98	40.43

Data normalization is implemented before the method is applied to the data itself. This action is intended to avoid bias of variables that have greater values compared to others, which is a normal behavior of the data.

**Table 6: RFE with normalization**

<b>Time shift (hh:mm)</b>	<b>Data ratio (%)</b>	<b>Average error (%)</b>
00:30	4.98	38.43
01:00	4.98	18.60
02:00	4.98	30.95
06:00	4.98	21.53

We expected better results when using the RFE method with normalization, because the RFE criterion could be biased in such a way that greater values would have more influence than smaller ones. This fact was observed in the results of Table 5 and Table 6.

### **Experiment 6 - Pearson Correlation**

Different numbers of variables, considered as relevant ones, were chosen for use in each test, as can be seen by the results.

**Table 7: Pearson Correlation**

<b>Data ratio (%)</b>	<b>Time shift (hh:mm)</b>	<b>Average error (%)</b>
2.49	00:30	22.47
	01:00	18.50
	02:00	14.30
	06:00	22.61
3.74	00:30	24.90
	01:00	19.06
	02:00	15.65
	06:00	24.27
4.98	00:30	17.51
	01:00	15.17
	02:00	15.34
	06:00	24.11
8.72	00:30	26.75
	01:00	17.92
	02:00	28.74
12.45	00:30	21.57
	01:00	25.28
	02:00	30.72

Using this method we obtained the prediction models with best accuracy. Table 7 presents the three best models obtained in all experiments.

## 5 Conclusion

The Vetta Artificial Intelligence Framework can be applied successfully to make on-line prediction of silicon grade in pig iron using only data of supervisory and laboratory databases. The full integration of several artificial intelligence algorithms and dimensionality reduction techniques makes the system capable of managing raw data obtained directly from the supervisory system, no matter how frequently information are collected.

The accuracy of the system could be improved drastically if we trained it for periods of normal functioning of the blast furnace. In the experiments, the period used to train and validate prediction models was a period of anomalous functioning. Even for such a period, we reached 85% accuracy. For periods of normal functioning we expect to reach an accuracy of 95% or greater, based on historical data we have for problems with similar levels of anomalous behavior in the training samples.

## 6 References

- [1] CARVALHO, B.P.R. **A study of methods and metrics for dimensionality reduction**. Technical report VT-2003/01, Vetta Technologies, 2003
- [2] DAVIS L. **Handbook of Genetic Algorithms**. Van Nostrand, New York, 1991
- [3] GUYON I., WESTON J., BARNHILL S., VAPNIK V. **Gene Selection for Cancer Classification using Support Vector Machines**. Machine Learning, 2000
- [4] KENNEY J., KEEPING E. **Mathematics of Statistics**. Linear Regression and Correlation, Van Nostrand, 1962
- [5] OJA, E. **PCA, ICA, and Nonlinear Hebbian Learning**. Helsinki University of Technology, 1995
- [6] PARTRIDGE M. G., CALVO R.A. **Fast dimensionality reduction and simple PCA**. Intelligent Data Analysis, 1998
- [7] PELIKAN, M. GOLDBERG, D.E., CANTÚ-PAZ, E. **BOA: The bayesian optimization algorithm**. Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-99, 1999
- [8] RUMELHART D. E., HINTON G.E., WILLIAMS R.J. **Learning internal representations by error propagation**. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, MIT Press, Cambridge, 1986
- [9] VAPNIK, V. **The nature of statistical learning theory**. Springer, NY, 1995
- [10] VAPNIK, V. **Statistical learning theory**. Wiley, 1998

# PREDICTION OF SILICON GRADE IN THE PIG IRON USING ARTIFICIAL INTELLIGENCE TECHNIQUES

Bernardo Penna Resende de Carvalho<sup>(1)</sup>  
André Luiz de Senna<sup>(2)</sup>  
Fábio Mayrink Morais<sup>(3)</sup>

## ABSTRACT:

In this work we present a system based on techniques of Artificial Intelligence (AI) combined in an unpublished way in literature to predict the silicon grade of samples of pig iron from blast furnace 2 runners in Acesita. Silicon grade is one of the main indicators of pig iron quality, and therefore a system capable of making accurate predictions allows the furnace manager to take corrective action more quickly, thus improving the process quality.

The system uses a combination of machine learning, neural networks and bayesian optimization algorithm, and allows online prediction of the silicon grade in the next run of pig iron, using furnace and process information, in addition to the results of mineralogic and chemical analyses of raw materials. The models are exclusively constructed on the basis of historical information from the supervisory and laboratories databases.

The system operates with error ranges lower than 5% in normal periods and lower than 15% in periods with anomalies. Tools for relevant data selection are used, without user intervention, which eliminate noise and redundancy of the supervisory data. Such tools explore varied characteristics of the data, and when used together, they generate a consistent data set for use in the AI models. The models are constantly updated through online training evaluations, allowing the automatic incorporation of alterations in the characteristics of the furnace, the process, or the raw materials.

**KEYWORDS:** artificial intelligence, prediction, quality.

---

<sup>(1)</sup> BSc, Electric Engineer - UFMG, MSc, Electric Engineer – UFMG, Research and Development Analyst - A.I. Division - Vetta Technologies.

<sup>(2)</sup> BSc and MSc, Computer Science – UFMG, PhC, Computer Science - UFMG, CTO - Vetta Technologies.

<sup>(3)</sup> BSc, Physics - UFMG, MSc, Metalurgy and Mines Engineering– UFMG , Reduction Technical Manager – Acesita S.A.