Predicting Paper Making Defects On-line Using Data Mining

Robert Milne and Mike Drummond
Intelligent Applications Ltd
1 Michaelson Square, Livingston
West Lothian, EH54 7DP
Scotland
UNITED KINGDOM

Patrick Renoux
Caledonian Paper plc
Meadowhead Road,
Irvine, KA11 5AT
Scotland
UNITED KINGDOM

Abstract: This paper describes an application that was jointly developed by *Caledonian Paper* and *Intelligent Applications Ltd* for the early prediction of paper defects from process data, so that corrective action can be applied before the defect becomes too significant. Correlations between process data and past faults were extracted and then programmed into an on-line predictive software model which is able to analyse current process data in real time, looking for bad patterns which may lead to defects in the paper. Depending on the degree of severity of defect that the model predicts, and the nature of the developing problem, the machine operators can take steps to prevent the defect from becoming so significant as to result in salvage. This article describes the way the application was developed and shows how data mining can be successfully applied to the paper industry.

Dr Robert Milne is the founder and Managing Director of Intelligent Applications Ltd. He is considered one of the most knowledgeable persons in the UK with regard to expert system applications in manufacturing. He has a BSc. in Electrical Engineering and Computer Science with special emphasis on Artificial Intelligence from the Massachusetts Institute of Technology (1978) and a Ph.D in Artificial Intelligence from Edinburgh University (1981). Dr. Milne has over 75 technical publications. He is a member of the American Association of Artificial Intelligence, the ACM, AISB and the British Computer Society. He is currently on the National Committee for the British Computer Society Specialist Group on Expert Systems. He was Conference Chairman of Expert Systems 1991, 1992 and 1996, the UK national expert systems conference. His a Chartered Engineer and a Certified European Engineer.

Dr Mike Drummond is a Software Engineer with Intelligent Applications Ltd. He has a BSc. (1979) and a Ph.D (1982) in Natural Philosophy from Glasgow University. He has been working in the area of applying expert systems to industrial applications since 1988.

Mr Patrick Renoux is a Graduate Engineer with Caledonian Paper plc. He has a Diplôme d'Ingénieur degree with a specialisation in Industrial Automatic and Control Systems from the Université Technique de Compiegne. He is responsible for the development of the models involved in the Prefault system and the investigation of other AI related techniques applicable to the paper industry. This work is part of on-going development by Caledonian Paper to improve the paper production process and increase the quality of the paper delivered to the customers.

1. Overview of the Project

Discussions began a few years ago between Caledonian Paper (CP) and Intelligent Applications (IAL) and focused on predicting paper salvage using process data. The two companies had worked together before on the development of a software tool to diagnose drive faults in the mill [Milne and Nelson, 1995]. The underlying approach had been to build up a library of past faults, and then match new fault symptoms into the library to find the most similar past one using the technology of case -based reasoning (CBR) [Kolodner, 1993]. Discussions took place between CP and IAL to investigate whether a similar technical approach could be applied to predicting salvage from process data. This case study describes what followed on from those discussions. This work received financial support from Ayrshire Enterprise.

CP, part of the UPM Group, is one of the UK's most modern paper mills, and is based on a 50 hectare site in Irvine, Scotland. With investment in excess of £200 million, it produced its first reel of Light Weight Coated (LWC) paper in April 1989. Since then, the company has become established as a leading supplier of LWC to publishers and printers throughout the world with an annual output of over 200,000 tonnes.

IAL are a privately owned UK company, specialising in the application of advanced software techniques to industry with a particular focus on diagnostic systems. They developed Amethyst for vibration analysis which is believed to be the most widely used expert system based tool in maintenance world-wide. They also received the Queens Award for Technological Innovation for Amethyst. They are currently working on advanced modelling techniques for turbine diagnostics, and on applications of data mining.

comparing current process data with historical examples of process data for which the quality of the final customer reel was known. However,

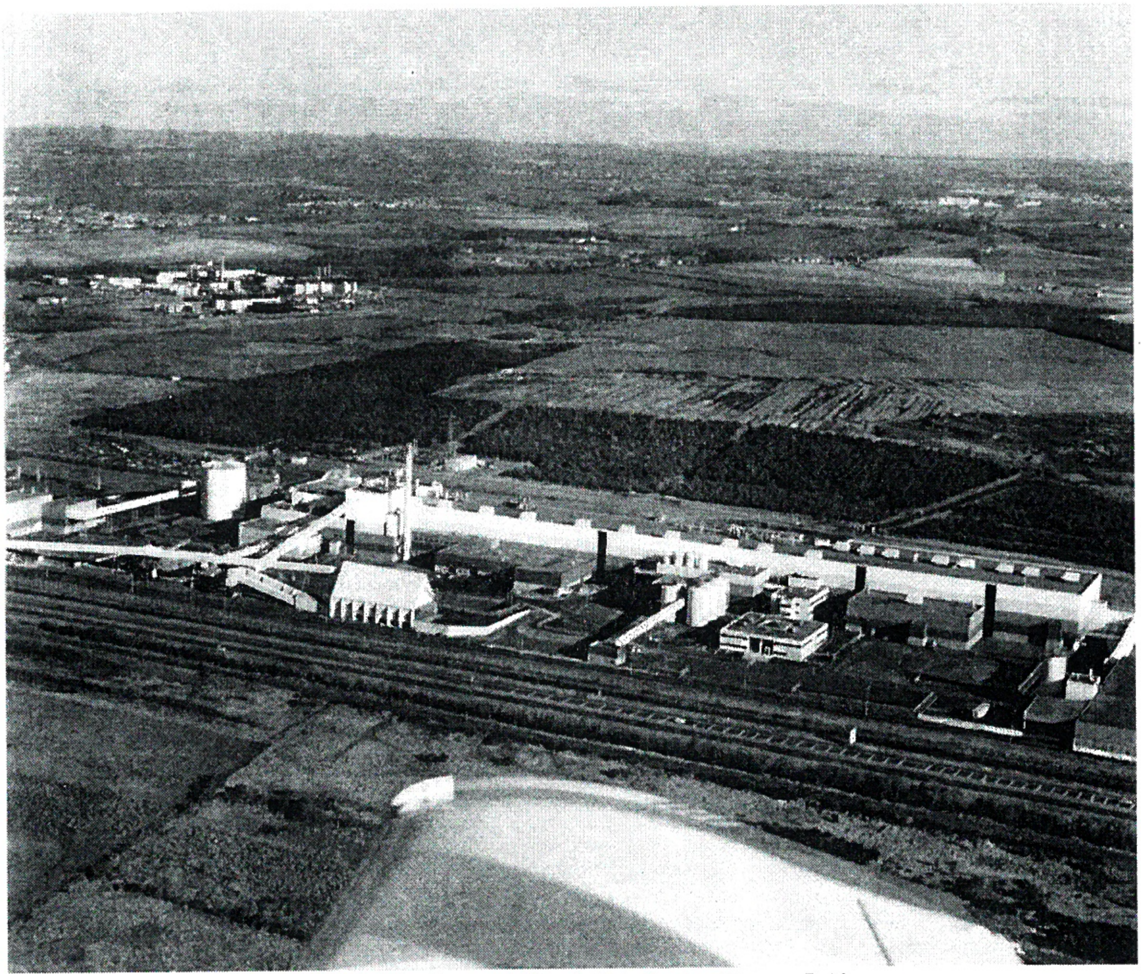

Figure 1. The Paper Mill at Caledonian Paper, Irvine

CP has a very small percentage of salvage in any given time period. One of their priorities is to reduce the salvage to as low a level as possible [Nelson and Renoux, 1997]. For CP, the main paper defect is deviation. This is when the customer reels build up with mis-aligned edges and therefore veer off to the side resulting in uneven edges. Corrugation also occurs occasionally, and is seen as a barring effect with finger like rings round the circumference. Whilst there are many possible reasons for paper defects, many of which are not related to the profile information directly, it was thought at the start of this work that a significant proportion of defects could be related back to profile data.

2. Data Mining for Predicting Defects

The underlying technical approach behind this work for predicting salvage is based on due to the volume of data to be analysed and the number of discrete data points to be compared, it was quickly recognised that this would be very difficult using conventional database techniques. An alternative to conventional data analysis techniques was to apply data mining to the problem [Fayyad et. al, 1995].

Data mining is a set of techniques for searching through data sets, looking for hidden correlations and trends which are, for all intents and purposes, inaccessible using conventional data analysis techniques [Parsaye and Chignell, 1993]. IAL have completed a range of data mining projects in other areas such as marketing, predicting competitors bid prices in the power industry and predicting the loyalty of customers to mortgage lenders. Although these applications on the surface are very different to the prediction of paper defects, in reality there are actually many common issues with predicting paper salvage. Both had large amounts of historical data. In both situations, the

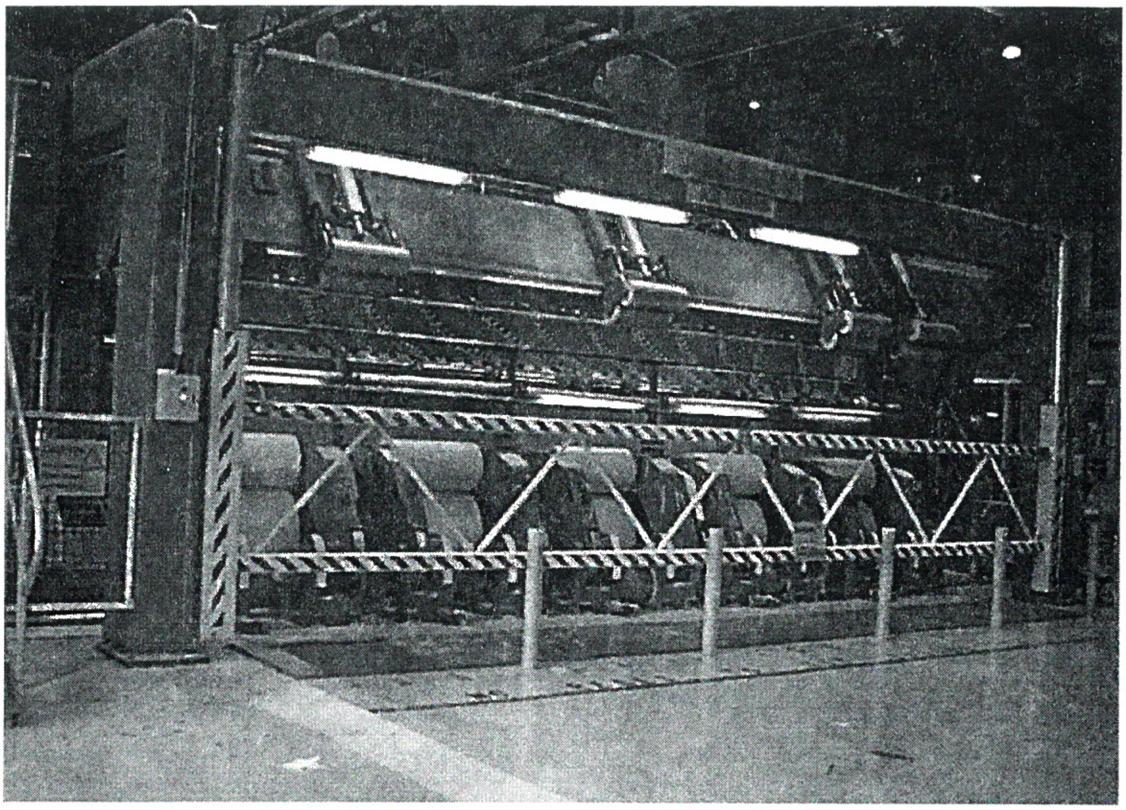

Figure 2. The winder station, where the paper reels are split into individual slices to match customer requirements. This is where the faults arise that are covered by this system

requirements were to model a very dynamic and complex process. In addition, IAL have developed many diagnostic systems for manufacturing, and their background expertise in this area meant that they were well placed to provide support to CP during this work.

For data mining to be successful, it is normally necessary to have at least several hundred historical examples of data records corresponding to each outcome of interest [Frawley et al, 1992]. In this case, CP had many thousands of historical examples of good paper reels, and enough historical examples of defective reels to form a good basis for applying data mining. Sometimes it can be very difficult to obtain access to the historical data due to a range of possible technical reasons. This can happen when customers data sets are distributed across many different types of databases since this can make it difficult to extract a complete set of process data corresponding to a single paper reel. However, CP have invested heavily in their IT infrastructure and were well placed to provide the level of data access which was needed for this work.

The business of data mining is growing for many reasons, including the fact that most business organisations are building up vast amounts of data concerning their manufacturing processes, their customers, their markets, and other important items of data. It is estimated that the volume of stored data in the world doubles every 18 months. One of the motivations behind storing these vast amounts of data is that it is thought to contain useful business information which describe past performance, or decisions. It should therefore be possible to use this historical information to improve future business or process decisions. However, a real problem in doing this is that the sheer volume of the data can often obscure the key information contained within the data. For many organisations, this makes the process of manual data analysis supported by perhaps some business graphics and statistical tools very difficult and cumbersome, resulting in information which is less than that which is potentially available from the historical data. This is where data mining comes in. Data mining is a set of techniques taken from the areas of machine learning and artificial intelligence which address the data analysis problem in a much more effective way by semi-automatically searching through the data sets and extracting useful trends and inter-relationships which would be extremely difficult to do using other more manual techniques.

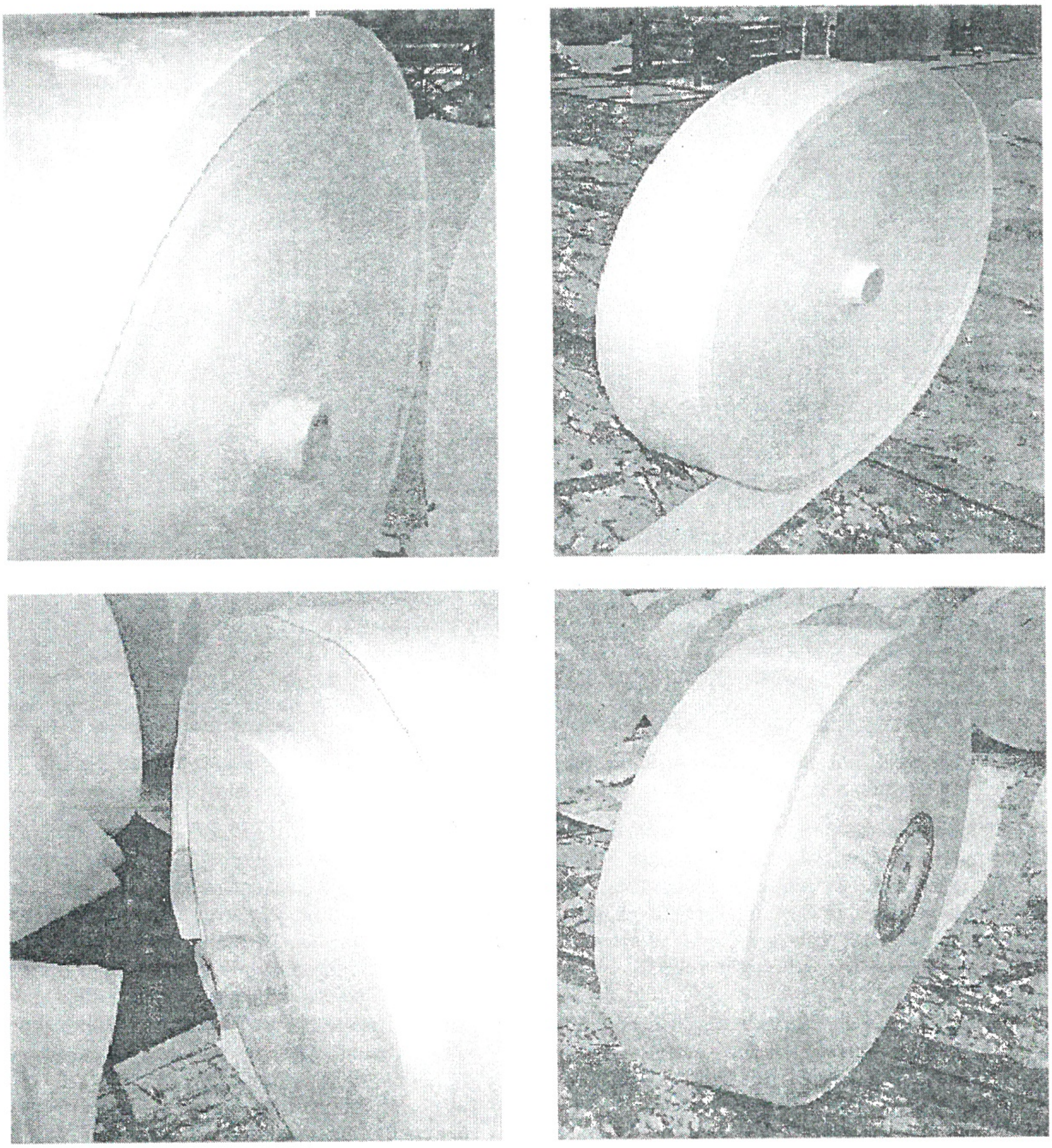

Figure 3. Some Examples of Wound Customer Reels Showing Deviation

There is a small number of core techniques which support data mining. These include neural networks, statistical clustering, CHAID analysis and induction. It was decided to use induction for this work since it is one of the few approaches which results in a human understandable software model. Most of the other approaches generate, in effect, a black box which is not able to provide any reasoning behind its findings. Induction, on the other hand, results in a decision tree which can be read and understood by a person. This has a significant impact on development times since the developers can more quickly understand the combination of attributes within the process data which the data mining has identified as being relevant for predicting paper defects. A further benefit of using this approach is that sometimes the analysis will identify some correlation in the data which were valid historically for relating paper defects to process data, but which are of no value to make predictions in the future. Using neural nets or some of the other approaches, it would have been more difficult to identify these loop-holes in the data leading to an unduly optimistic view of the resulting predictive model.

3. The Process Data

CP save the manufacturing data into their quality system database using a range of integrated data acquisition systems. At various stations in the mill, the paper properties are measured by gauges which continually traverse the moving paper web. These properties include basis weight, moisture content and caliper. Basis weight is the weight per unit area of a mass of paper. It relates to the mathematical product of the density and thickness of the paper being measured. Units of measurement are grams per square meter (gsm). The scanner uses a radioactive source (Kr85) of beta rays and a detector/receiver to measure basis weight. Moisture represents the percentage of moisture content in the paper. Moisture measurement is carried out by an infra-red absorption method which uses the fact that water molecules absorb electromagnetic energy in different bands. Caliper represents the sheet thickness in microns. The sensor measures the thickness of a moving sheet by measuring the magnetic reluctance between upper and lower contacts of the sensors as they independently follow sheet contours.

The Supercalender scanner carries out a large number of scans in the cross direction for a full reel, taking 72 measurement points across the reel which is approximately 8 metres wide. For this application, the data collected from the scanner were averaged along the length of the reel and resulted in a set of 72 data points for each of caliper, basis weight and moisture.

The data from each reel could not be compared with the data from all other reels due to the different grades of paper which are produced in the mill. It was necessary to group the historical data based on bands of paper grade to ensure that comparisons between good and defective reels were valid.

4. Model Development

Before work began on the development of the predictive model, time was spent performing data visualisation to compare examples of good and bad profiles. Initially, this work was done using data which had been averaged along the entire length of the reels. It was clear that a lot of detail would be lost from the data by doing this, but the IT infrastructure that was in place when the project started meant that this was the most sensible place to start. Several months were spent doing visual and numeric comparisons between good and bad profiles, and although it was possible to deduce certain things about the differences based on the shape of the profiles, it was felt that the lack of a more precise method of comparing the profiles was limiting the work.

The next stage was therefore to obtain a data extract and import it into a data mining software tool to see what data mining could offer. Using the same data as had been used previously for the visual comparison work, work began on data mining using the ReMind software shell from CSI, a US company known for its work in case based reasoning. ReMind is an MS Windows based tool which supports induction for data mining. It was chosen partly because of the previous successful work done by Intelligent Applications using ReMind's inductive engine, and also because of its unique capabilities to support the programming of process and manufacturing knowledge to help guide the induction process. In effect, ReMind allows its users to enter qualitative information about the process being modelled (in this case the paper making process) to guide the induction. Based on the previous work done by IAL, this capability in the development tool can often make a significant difference to the accuracy of the resulting model.

Induction is straightforward to understand at the conceptual level [Quinlan, 1993]. In this context, induction analyses a number of examples of good and bad profiles, trying to determine which data attribute is most different between the good and bad examples. For example, it might determine that the difference in caliper between the edge of the reel and the next 10 adjacent data points show the largest degree of deviation between good and bad reels. Induction would then split the data into two groups based on that attribute. It would then analyse both of these two groups and again try to find out which attribute is most significant within the sub-groups at separating out the good from the bad profiles. Again, each group would then be split further. This process continues until there are many groups, or clusters, which are deemed to be an optimal set of groupings. This is analogous to performing market segmentation for a bank, for example. The bank may have 3 million mortgage customers, but it knows that there are certain natural groupings of customers, within which there is a certain degree of similarity between customers.

The inductive process itself, once completed, can be represented by a decision tree, with splits in the tree being based on the values of attributes in the process data [Riesback and Schank, 1989]. A partial example of such a tree is shown below.

reel. Additionally, the significance of any given attribute may not always be constant. A particular attribute may be more important in certain situations compared to others. Therefore

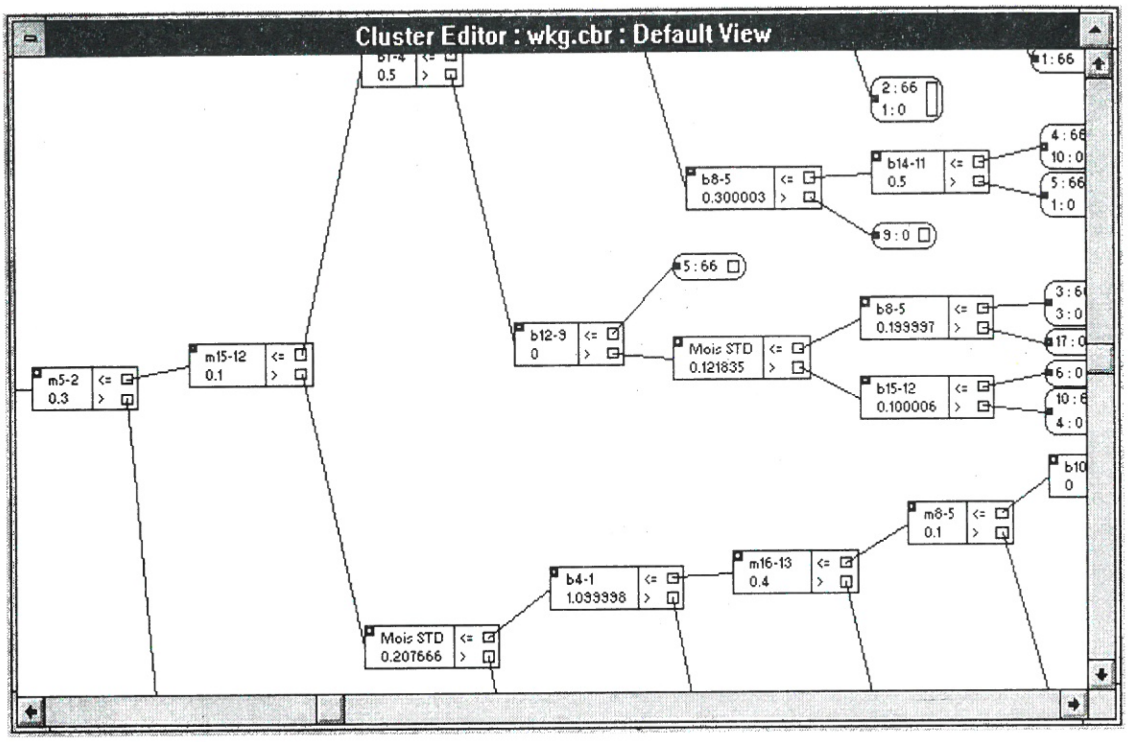

Figure 4. The Decision Tree Produced by the Induction Process Using ReMind

the significance of each attribute is dependent on the values of other attributes in the process data. The data mining technique which was used for this work is able to identify these contextual dependencies and generates different significance figures for each attribute based on the values of other attributes in the data.

Looking at the left-hand side of the diagram, there is a data field called 'm5-2' in the process data. If the actual value of this field in a new reel is less than or equal to 0.3, then the next data item to check is 'm15-12'. If the actual value in the new reel is greater than 0.1, then the next data item to check is 'Mois STD'. This process of walking through the decision tree continues until the end of a branch is reached. These 'leaves' of the tree contain a group of historical reels which are predominantly either good or defective. The actual prediction for a new reel is based on the probability distribution within the final leaf in the tree that the prediction process ends up at.

The diagram shows how induction uses various attributes in the data sets to discriminate between likely good and defective reels, and these are indicated in the boxes in the diagram. The decisions are actually based on whether the value of an attribute in a new reel is above or below a threshold value appropriate for that attribute. The choice of what the thresholds should be tested against are automatically determined by the data mining software. This approach has the benefit that it automatically determines the relative significance of each attribute in the process data based on the impact that each attribute has on the quality of the final

Once completed, this tree can be used to classify new paper reels based on the attributes that the tree is built on. When the model goes on-line, the process data are analysed by the model by percolating them down through the tree as the paper is being made, and depending on which sub-group it ends up in, the model will indicate whether a paper defect is likely to develop or not. A further benefit of using induction for this work was that, when the final predictive model goes on-line, it can not only generate a prediction concerning likely defects, but it also generates the probability of the prediction being correct. The probability is based on the probability distribution of good and bad reels in the final *leaf* that the model used to base its prediction on. This is very useful since it is possible to impose a threshold on the system so that only predictions which are more likely to be correct than, say, 85%, should be acted upon by the operators.

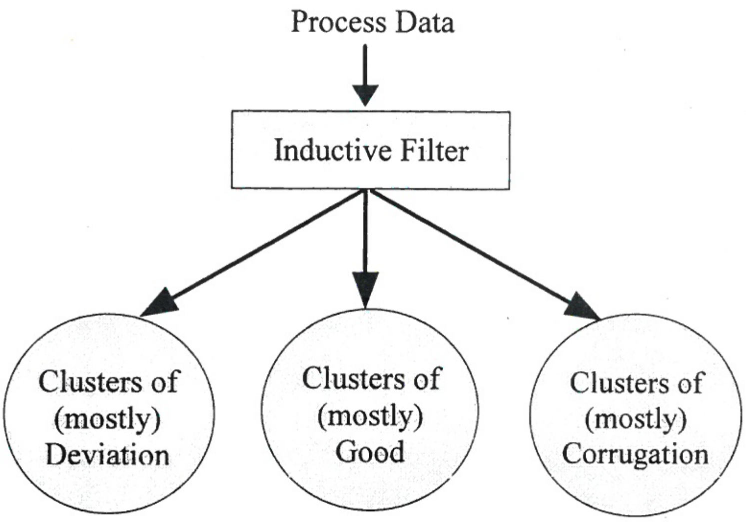

Figure 5. Showing how the inductive filter segments the process data from many different paper reels into groups. Each of these groups has an outcome which determines whether a reel is predicted to be good, deviated or corrugated

The ReMind software has specific tools which allow the inductive process to be fine-tuned based on process and manufacturing knowledge. This can be very helpful where such knowledge exists. In this case, it was suspected that deviations in the profile near the edge of the reels appeared to be very significant, and so this information was used to guide and provide an initial focus for the inductive engine in ReMind.

The data had to be pre-processed in order to present the data mining software with an ASCII text file which was the only format recognised by the Development Environment of ReMind. This was done using MS Access, although Lotus 123 or MS Excel would also be suitable. A range of pre-processed attributes were extracted from the data for each customer reel including standard deviation, kurtosis and other statistics which were thought to have a bearing on what makes a bad profile. This is almost always very important for data mining work since, unless an infinite number of historical examples is available for the data mining tool, it may not produce the best model possible. This process of using process knowledge and expertise to support induction leads to a methodology known as Knowledge-Guided Induction. It is a blend of automatic machine learning techniques and human expertise about the domain being analysed, and represents a significant step forward over normal induction, usually resulting in more accurate predictive models than might exist using other automatic data mining techniques.

The process of building the model is an iterative one where an initial model is built using a sub-set of historical examples, and then it is tested using some of the other historical examples. By comparing what the model predicts would happen with what actually did happen with the historical test cases, it was possible to get a clear picture of the accuracy of the model. Additionally, it was possible to work out whether the errors were more or less equally distributed through each branch of the decision tree, or whether certain 'branches' in the tree were responsible for a disproportionate number of mispredictions. Once each branch was shown to be reasonably robust and correct, it would be frozen and work would begin on one of the other less accurate branches. This process iterates round a few times since making changes to one branch to improve it can result in an increased error rate for other branches which had previously been validated. This is where the bulk of the data mining effort was applied, and is still on-going. A key to this development process is recognising that at some point, enough of the branches will be accurate enough to enable the model to go live, and that if certain branches are not at the same level of accuracy, then any prediction which uses these branches can be either ignored by the operators, or displayed with an appropriate message to indicate that the operator should investigate further before taking corrective action based on the prediction.

At the time of writing, the model is receiving further development effort to improve its accuracy, but it has produced the following levels of accuracy.

	Correctly Predicted	Wrongly Predicted
Good Reels	92%	8%
Defective Reels	31%	69%

This means that 92% of the good reels were correctly predicted and 31% of the faulty reels correctly predicted. According to Roy McHattie, an Electrical/Instrument Superintendent at the mill *"These figures are very good given the quality of the data which is being processed. It is interesting to note that there are certain aspects of the predictive model which are counter-intuitive in that they indicate a likely defect based on data values which are not, on the surface, related to defects. Some of these relationships are particularly difficult to understand but this can be useful since it can encourage a re-think of some of their previously held beliefs about the process itself."*

5. Putting the Predictive Model On-line

The management at the mill made the decision last year that the model was showing enough promise to commence the development of an interface between the model and the manufacturing process itself [Milne and Nelson, 1997]. This was also a joint development effort between CP and IAL since work was required to extract the process data from the mills quality system databases and download the process data to the PC running the model. Working together, the two companies defined the technical interface between the mills VAX and the model which runs on a PC. CP developed the data download program which acts as the data feed to the PC where the predictive model runs. IAL developed software to receive the process data and then feed them through the predictive model. They also developed an appropriate user interface which shows the operators which profile is currently being processed at any given time, what the model is predicting for that profile and what the probability of it being correct is. The operator is able to enter corrective actions which can be applied in an attempt to prevent the defect from developing further. This is necessary for the on-going monitoring of the model. Since the manufacturing process is dynamic, the accuracy of the model may degrade over time. By tracking the accuracy of the models predictions, this can be checked and the model refined to compensate for any changes in the manufacturing process.

On a commercial note, the costs for putting the model on-line exceeded the costs of the model development. According to IAL, this is not uncommon. During the model development stage, the data mining work is usually performed off-line working with data extracts. The technical issues related to interfacing the model in an on-line way depend on the existing IT systems that are in place already. These issues are to do with systems integration and are very different to the difficulties associated with the data mining task itself. However, the commercial risks associated with this type of work are minimised since the accuracy of the model is known before the larger costs of model integration are incurred. The work on the integration does not commence until the customer knows how accurate and therefore how cost-effective the model is going to be.

6. Current Project Status

Today, the interface has been completed, and the model has commenced trials of running on a live basis as a decision support tool. Work to improve the accuracy of the model is continuing. The accuracy of the model did decrease when the system went live. This was expected and was due to changes in the manufacturing process and also changes in the way that the data were measured between the time of the historical examples and the time the system went live.

When the system's credibility has been established in live operation, the intention is to provide it as a tool for operators.

7. Conclusion

The full benefit of using the system will be some time in the future, but more intermediate benefits are in sight. CP is close to having a system which will support their manufacturing goal of reducing the level of salvage. The strategic objective of continuous improvement within CP will never disappear, but the development of this system has done the difficult job of identifying how to improve on a process which already functions almost all of the time without error anyway.

Lessons have been learnt during this work. The model development has taken longer than was originally expected. The testing and validation of the many models which have been produced during the iterative development cycles took longer than expected. This is largely due to the volume of data that was required to perform an adequate level of validation. The need to adopt a meticulous approach to documenting model development and validation also became apparent since these tasks were so involved and it would have been easy to lose track of a few details found early on in the development process which proved very useful later.

This work is under way. CP have plans to look further back down the process towards the Paper Machine itself. (Currently the model is interfaced to the manufacturing process at the Supercalenders). It is more important that the model is accurate at the paper machine because, if the operator takes any corrective action based on what the model is telling him at the paper machine, it could have a very significant impact further down the line.

Finally, since the process data being analysed are based on an average of the entire length of the paper reels, work is commencing to look at using non-averaged process data. It is almost certain that a new model could be built using this non-averaged data which would result in significantly higher levels of accuracy than that which is currently possible.

REFERENCES

FAYYAD, U., PIATETSKY-SHAPIRO, G. and SMYTH, P., **From Knowledge Discovery to Data Mining: An Overview,** in U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (Eds.) Advances in Knowledge Discovery and Data Mining, AAAI/MIT PRESS, 1995.

FRAWLEY, W., PIATETSKY-SHAPIRO, G. and MATHEUS, C., **Knowledge Discovery in Databases: An Overview.** AI MAGAZINE, Autumn 1992.

KOLODNER, J., **Case Based Reasoning,** MORGAN KAUFMANN PUBLISHERS, 1993.

MILNE, R. and NELSON, C., **Electrical Drive Diagnosis Using Case Based Reasoning,** IEE Colloquium on CBR: Prospects for Applications, London, UK, 7 March 1995.

MILNE, R. and NELSON, C., **Using Data Mining To Learn To Be Able To Predict Paper Making Defects On-Line,** Unicom Seminar, CBR & Data Mining: Putting The Technology To Use, London, UK, 18-19 March 1997.

NELSON, C. and RENOUX, P., **Caledonian Plans To Reduce Deviant Reels By 10%,** PAPER TECHNOLOGY, February 1997, p.7.

PARSAYE, K. and CHIGNELL, M. , **Intelligent Database Tools and Applications,** JOHN WILEY, 1993.

QUINLAN, J.R., **C4.5 Programs for Machine Learning,** MORGAN KAUFMANN PUBLISHERS, CA, 1993.

RIESBACK, C. and SCHANK, R. , **Inside Case-Based Reasoning,** Hillsdale, N.J. Lawrence Erlbaum, 1989.